University of Pennsylvania
ScholarlyCommons

Publicly Accessible Penn Dissertations

2020

The Penetrance Of Pancreas Agenesis Caused By Gata6 Mutations Is Modified By A Non-Coding Snp

Siddharth Kishore
University of Pennsylvania

Follow this and additional works at: https://repository.upenn.edu/edissertations

 Part of the Cell Biology Commons, and the Developmental Biology Commons

Recommended Citation

Kishore, Siddharth, "The Penetrance Of Pancreas Agenesis Caused By Gata6 Mutations Is Modified By A Non-Coding Snp" (2020). *Publicly Accessible Penn Dissertations*. 4169.
https://repository.upenn.edu/edissertations/4169

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/edissertations/4169
For more information, please contact repository@pobox.upenn.edu.

The Penetrance Of Pancreas Agenesis Caused By Gata6 Mutations Is Modified By A Non-Coding Snp

Abstract

GATA6 is a critical regulator of pancreas development, with heterozygous mutations in this transcription factor being the most common cause of pancreas agenesis. However, patients harboring GATA6 mutations exhibit variability in disease phenotypes. We have used a pancreatic agenesis patient-induced pluripotent stem cell model to study this disorder. We found that after correcting the coding mutation in a pancreas agenesis patient's iPS cell line, GATA6 protein expression was comparable to other wild type stem cell lines at the definitive endoderm stage of development but still depressed in pancreatic progenitors. To investigate this finding, we screened the regulatory regions of the GATA6 gene and identified a SNP in a 3' regulatory region of GATA6, with the patient carrying the minor allele variant. We tested the SNP in 32 further patients with pancreatic agenesis caused by GATA6 mutations and found that the frequency of the minor allele was enriched in the pancreatic agenesis cohort. The minor allele variant disrupted binding of the orphan nuclear receptor RORα and reduced GATA6 expression and efficiency of pancreas differentiation. Our work highlights a possible genetic modifier contributing to the pancreatic agenesis phenotype in patients with GATA6 mutations and shows the benefits of using pluripotent stem cells to study the effects of non-coding genetic variants in modifying disease penetrance.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Cell & Molecular Biology

First Advisor

Paul Gadue

Subject Categories

Cell Biology | Developmental Biology

This dissertation is available at ScholarlyCommons: https://repository.upenn.edu/edissertations/4169

THE PENETRANCE OF PANCREAS AGENESIS CASUED BY GATA6 MUTATIONS IS MODIFIED BY A NON-CODING SNP

Siddharth Kishore

A DISSERTATION

in

Cell and Molecular Biology

Presented to the Faculties of the University of Pennsylvania

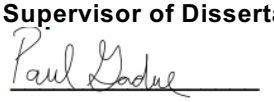
in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2020

Supervisor of Dissertation



Dr. Paul Gadue, Associate Professor of Pathology and Laboratory Medicine

Graduate Group Chairperson

Dr. Daniel S. Kessler, Associate Professor of Cell and Developmental Biology

Dissertation Committee

Dr. Patrick Seale, Associate Professor of Cell and Developmental Biology

Dr. Doris Stoffers, Sylvan H. Eisman Professor of Medicine

Dr. Daniel Kessler, Associate Professor of Cell and Developmental Biology

Dr. Ken Zaret, Joseph Leidy Professor of Cell and Developmental Biology

THE PENETRANCE OF PANCREAS AGENESIS CASUED BY GATA6 MUTATIONS IS MODIFIED BY A NON-CODING SNP

COPYRIGHT

2020

Siddharth Kishore

ABSTRACT

THE PENETRANCE OF PANCREAS AGENESIS CASUED BY GATA6 MUTATIONS IS MODIFIED BY A NON-CODING SNP

Siddharth Kishore

Paul Gadue

GATA6 is a critical regulator of pancreas development, with heterozygous mutations in this transcription factor being the most common cause of pancreas agenesis. However, patients harboring *GATA6* mutations exhibit variability in disease phenotypes. We have used a pancreatic agenesis patient-induced pluripotent stem cell model to study this disorder. We found that after correcting the coding mutation in a pancreas agenesis patient's iPS cell line, GATA6 protein expression was comparable to other wild type stem cell lines at the definitive endoderm stage of development but still depressed in pancreatic progenitors. To investigate this finding, we screened the regulatory regions of the *GATA6* gene and identified a SNP in a 3' regulatory region of *GATA6*, with the patient carrying the minor allele variant. We tested the SNP in 32 further patients with pancreatic agenesis caused by *GATA6* mutations and found that the frequency of the minor allele was enriched in the pancreatic agenesis cohort. The minor allele variant disrupted binding of the orphan nuclear receptor RORα and reduced GATA6 expression and efficiency of pancreas differentiation. Our work highlights a possible genetic modifier contributing to the pancreatic agenesis phenotype in patients with *GATA6* mutations and shows the benefits of using pluripotent stem cells to study the effects of non-coding genetic variants in modifying disease penetrance.

Table of Contents

LIST OF TABLES

LIST OF ILLUSTRATIONS

CHAPTER 1: INTRODUCTION AND OVERVIEW

1.1 The pancreas

The human pancreas is an organ that develops in the foregut endoderm and originates from two separate primordia, the dorsal and ventral buds that arise from either side of the distal foregut endoderm. The pancreas is made up of a variety of distinct cell types such as exocrine, ductal and endocrine cells (Figure 1.1). The exocrine cells, which constitute a majority of the pancreatic tissue, secrete digestive enzymes such as amylase, proteases, lipases and nucleases to the duodenum via the pancreatic duct that runs through the middle of the pancreas. These enzymes catalyze the breakdown of lipids, proteins and carbohydrates and play an important role in nutrient digestion (Shih, Wang and Sander, 2013). The endocrine cells are scattered throughout the pancreas and located in small tight clusters in structures known as the Islets of Langerhans. Within the islets, vasculature, neurons and mesodermal derived stromal cells ensure the proper functioning of the endocrine cells (Pan and Wright, 2011). The functioning of endocrine cells are mostly independent of the exocrine pancreas. Each type of endocrine cell within the islet produces a specific type of hormone. α-cells secrete glucagon, β-cells secrete insulin, δ-cells secrete somatostatin, ε-cells secrete ghrelin and γ-cells secrete pancreatic polypeptide (PP) (Da Silva Xavier, 2018). Glucose homeostasis is maintained in the body by the α-cells, β-cells and δ-cells. α-cells secrete glucagon when blood glucose is low and β-cells secrete insulin when blood glucose is high. δ-cells act as negative regulators of α-cells and β-cells. δ-cells, ε-cells and γ-cells function to regulate nutrient metabolism. Human islets are composed of 55% β-cells, 35% α-cells and 10% δ-cells with all cell types intermixed within the islet architecture. The total number of

islets in an adult human pancreas can vary between 300,000 and 1,500,000 (Brissova, Fowler, *et al.*, 2005).

Figure 1.1: Anatomy of the human pancreas

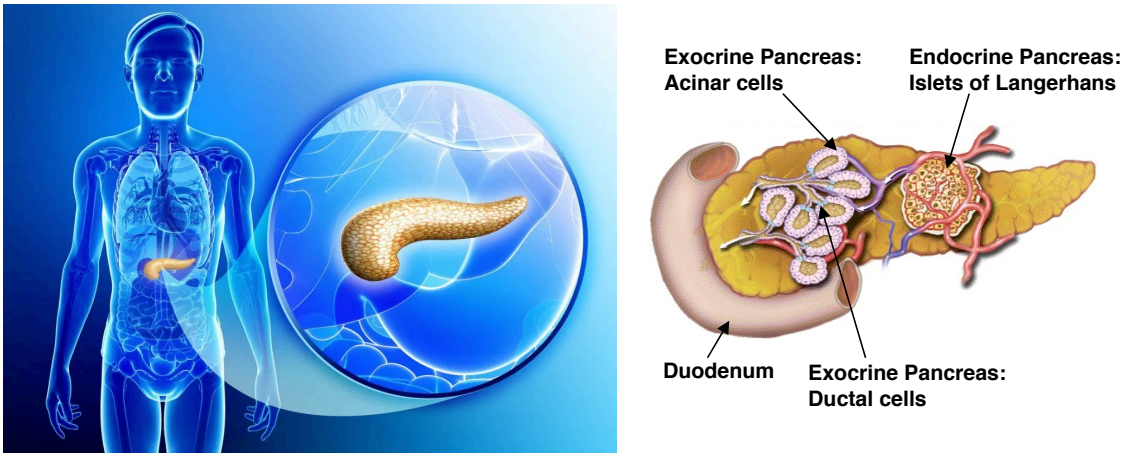


Figure 1.1: Anatomy of the human pancreas

The human pancreas is about 6 inches long and sits across the back of the abdomen, behind the stomach. The head of the pancreas is on the right side of the abdomen and is connected to the duodenum through a small tube called the pancreatic duct. The pancreas is comprised of exocrine and endocrine cells. The exocrine pancreas makes and secretes digestive enzymes into the duodenum. This includes acinar and duct cells with associated connective tissue, vessels, and nerves. The exocrine components comprise more than 95% of the pancreatic mass. The endocrine cells in the pancreas are the islets that make and secrete insulin, glucagon, somatostatin and pancreatic polypeptide into the blood. The islets comprise 1-2% of pancreatic mass. Adapted from Blausen.com staff (2014).

1.1.1 Endocrine pancreas

Given the key role of pancreatic endocrine cells in maintaining glucose homeostasis and in diseases such as diabetes, understanding how these cells develop and function is very important. The endocrine population of the pancreas is organized into small

structures or mini-organs called islets of Langerhans. There are five types of endocrine cells within the islet that produce specific type of hormones. α-cells secrete glucagon, β-cells secrete insulin, δ-cells secrete somatostatin, ε-cells secrete ghrelin and γ-cells secrete pancreatic polypeptide (PP) (Da Silva Xavier, 2018). The average size of an islet is 108μM and the total islet mass represents 1 to 2% of the pancreas (Ionescu-Tirgoviste *et al.*, 2015; Da Silva Xavier, 2018). The islets are highly vascularized structures that have five times more capillaries than exocrine pancreatic tissue and receive 5-15% of the entire pancreatic blood flow even though they consist of only 1-2% of the pancreatic mass (Ballian and Brunicardi, 2007). Islet architecture varies between species with striking differences seen when comparing human and mouse islets. In the mouse, the inner core of the islet is mainly β-cells and is surrounded by α-cells and δ- cells, while in humans the cells are scattered throughout the islet. Another key difference is in the proportion of α-cells to β-cells between mouse and human. The average cell composition of mouse islets is 77% β-cells and 18% α-cells whereas in humans it is 55% β-cells and 38% α-cells (Cabrera *et al.*, 2006; Dolensek, Rupnik and Stozer, 2015). These differences in islet structure between species have an impact on islet physiology and function, specifically related to calcium flux, metabolism and mitochondrial function (MacDonald *et al.*, 2011; Gregg *et al.*, 2016).

The transcription factor Neurogenin3 (NGN3) plays an essential role in endocrine development since it influences the expression of a diverse number of well-known transcription factors involved in pancreas development and β-cell function (PDX1, NKX6.1, SOX9, NKX2.2, NEUROD1, PAX4, RFX6, PAX6, GLIS3 and others) (Gasa *et al.*, 2004). NGN3 is a master regulator that specifies all 5 cell types of the endocrine

pancreas (Habener, Kemp and Thomas, 2005; Murtaugh, 2007; Sheets *et al.*, 2018). While the development of individual endocrine cell populations from Ngn3+ cells is still a matter of active research, it is known that counteracting transcriptional networks induce α-cells and β-cells. ARX and PAX4 are two transcription factors that repress each other and help define the α- versus β-cell populations. ARX drives the α-cell formation, and PAX4 is critical for β- and δ-cell formation (Collombat *et al.*, 2005; Kordowich *et al.*, 2012). Increasing our understanding of the extrinsic as well as intrinsic mechanisms that control these processes should facilitate the efforts to generate surrogate β-cells from human pluripotent stem cells (hPSCs), or to reactivate the function of important cell types within pancreatic islets that are lost in diabetes.

1.1.2 Exocrine pancreas

The exocrine pancreas is responsible for secretion of digestive enzymes, ions and water into the duodenum of the gastrointestinal tract. The digestive enzymes are essential for processing foodstuffs in meals to molecular constituents that can be absorbed across the gastrointestinal surface epithelium. Although enzymes from salivary glands, the stomach and the surface epithelium of the gastrointestinal tract also participate in the digestion of a meal, the exocrine pancreas plays a central and essential role in the digestive process. With loss of exocrine pancreatic function, absorption of nutrients in markedly compromised and malnutrition ensues (Pandol, 2010).

The secretion of ions and water is also an essential function for the exocrine pancreas. The flow of ions and water is necessary to transport the digestive enzymes from their origin in the pancreatic acinar cells to the intestine. In addition, the pH of the pancreatic secretions is alkaline due to a very high concentration of $NaHCO_3$ (up to 140 mM). At least one major function of the $NaHCO_3$ is to neutralize the acidic pH of the gastric chyme delivered to the intestine from the stomach. A neutral pH in the intestinal lumen is necessary for optimal function of digestive enzymes as well as gastrointestinal surface epithelial function (Pandol, 2010).

Anatomic studies demonstrate that the blood flow from the endocrine pancreas enters the capillaries of the exocrine tissue surrounding each of the islets before entering the general circulation (Ballian and Brunicardi, 2007). This “portal” system provides for the delivery of very high concentrations of hormones from the islets to the exocrine tissue surrounding the islets. The hormones from the islets include insulin, amylin, glucagon, somatostatin and PP. Although the full significance of the effects of these hormones on the exocrine pancreas is not known, the acinar cells that are involved in the regulation of digestive enzyme synthesis of the exocrine pancreas have insulin receptors (Pierzynowski *et al.*, 2018).

The signaling pathways underlying the development process include the Hedgehog system, the homeobox gene PDX1 and Notch signaling (Kim and MacDonald, 2002; Habener, Kemp and Thomas, 2005). Inhibition of Hedgehog signaling leads to ectopic budding of pancreatic structures in the stomach and duodenum (Parkin and Ingham, 2008). PDX1 expression in the gut tube during development marks the location of

pancreatic bud development. Notch signaling inhibits endocrine cell differentiation and promotes exocrine cell differentiation (Kim *et al.*, 2010).

1.2 Development of the human pancreas

Embryogenesis in humans lasts ~ 8 weeks post conception, after which the embryo is referred to as a fetus. Embryogenesis is marked by the specification of the 3 germ layers: ectoderm, mesoderm and endoderm. All adult tissues originate from these 3 germ layers. Based on a morphological scheme and staged by extension of time i.e. days post-coitum (dpc), human embryonic development is divided into 23 different Carnegie Stages (CS) (O'Rahilly and Muller, 2010). Using this classification, human pancreas development can be segmented into different Carnegie Stages, highlighted by the morphology of the pancreas as well as expression of key transcription factors (Jennings *et al.*, 2015a). Figure 1.2 highlights the signaling events and transcriptional networks that orchestrate the highly complex step-wise process of human pancreas development.

During early development, the region of the foregut endoderm from which the pancreas will arise is pre-patterned. Pancreas induction begins with the ventral and dorsal thickenings of the epithelial cells in the distal foregut, which occurs at CS9 (Piper *et al.*, 2004). During this stage, the definitive endoderm (DE) maintains communication with the visceral endoderm of the yolk sac (Jennings *et al.*, 2013). Following this, at CS10, endodermal folding specifies the foregut and hindgut. This folding restricts the opening

of the yolk sac to the midgut (Jennings *et al.*, 2013). The site of pancreas specification lies at the foregut-midgut boundary, which is known as the anterior intestinal portal (AIP).

Figure 1.2: Stages of human pancreas development

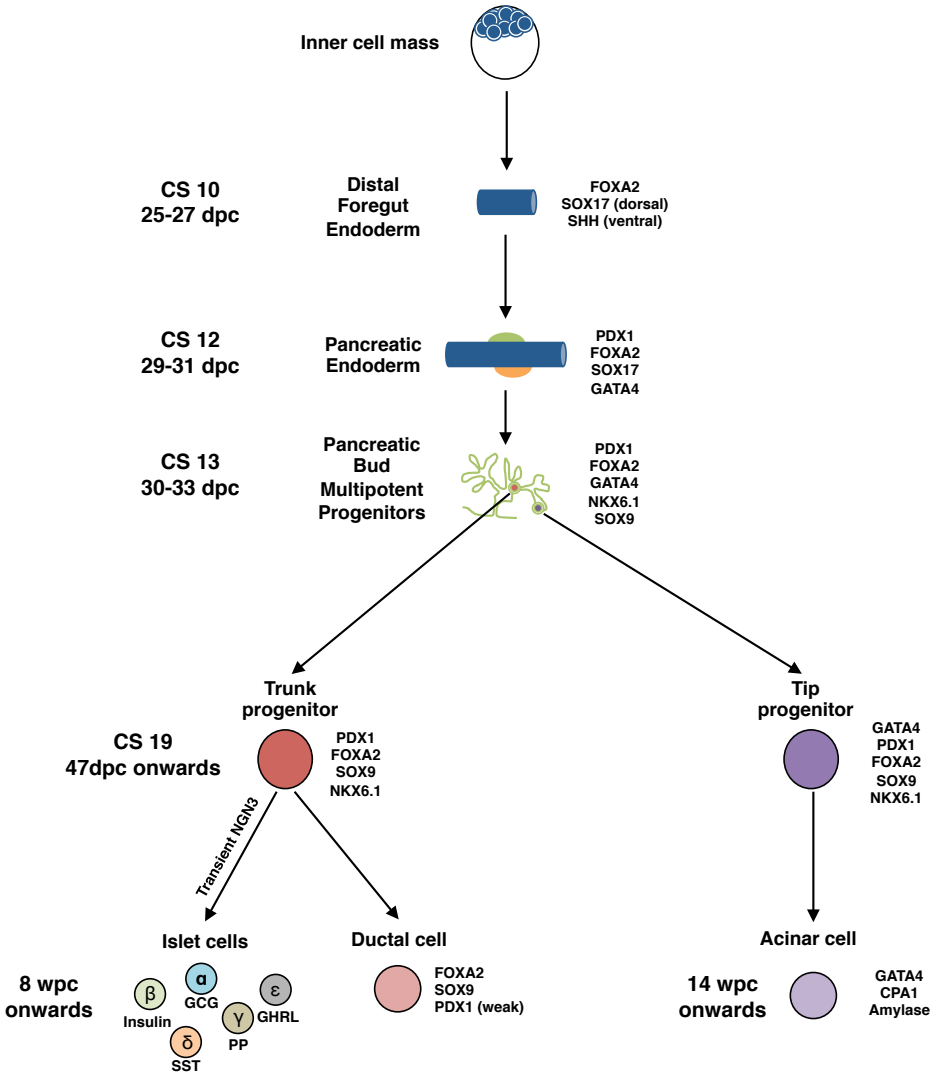


Figure 1.2: Stages of the human pancreas development

An illustration showing the timeline of human pancreas development from pluripotent stem cells in the inner cell mass of the blastocyst to fate committed islet, ductal and acinar cells. Key transcription factors that govern developmental stages and cell identity are highlighted. CS, Carnegie Stage; dpc, days post coitum; wpc, weeks post coitum.

The AIP has a lack of expression of sonic hedgehog and requires the expression of retinoic acid (Apelqvist, Ahlgren and Edlund, 1997; Kim and Melton, 1998; Hebrok, Kim and Jacques, 2000). The presumptive pancreatic endoderm is characterized by the expression of PDX1, PTF1A and SOX9. At this stage the pancreatic epithelium is tightly surrounded by mesenchyme, which expresses FGF10 at high levels. FGF10 regulates expression of PTF1A and SOX9 and is important for cell proliferation and expansion of the early pancreatic buds once they form (Pan and Wright, 2011; Shih, Wang and Sander, 2013). In humans, PDX1 is first detected at CS12, even though SHH could still be detected at CS10, which suggests a slightly later timing for the exclusion of SHH in humans (Jennings *et al.*, 2013).

The pancreas begins to develop by forming as two epithelial buds, a ventral and a dorsal pancreatic bud at opposite sides of the foregut endoderm by evagination into the surrounding mesenchyme. In both mouse and human, the dorsal bud forms first from the dorsal foregut endoderm. By CS13, the dorsal and ventral pancreatic buds are clearly defined and are marked by the transcription factors SOX9, PDX1, GATA4 and GATA6 (Piper *et al.*, 2004; Jennings *et al.*, 2013), all of which play a pivotal role in promoting human pancreatic growth (Stoffers *et al.*, 1997; Lango Allen *et al.*, 2012; Shaw-smith *et al.*, 2014). The emergence of the dorsal bud appears to be regulated by extrinsic cues from the surrounding vascular endothelial cells. This is followed by the evagination of the ventral pancreatic bud along with the liver from the ventral foregut endoderm. The ventral pancreatic bud receives cues form the cardiac mesoderm and the vitelline veins (Gittes, 2009; Pan and Wright, 2011). The human pancreas continues its expansion of proliferative progenitor cells for the remainder of the embryonic period. A unique aspect

of human pancreas development is the lack of early pancreatic endocrine differentiation as compared to mouse and chick embryos, which display significant early pro-endocrine patterning (Lammert, Cleaver and Melton, 2001; Bonal and Herrera, 2008; Villasenor *et al.*, 2010; Jennings *et al.*, 2013; Otter and Lammert, 2016). Another notable difference between the human and mouse pancreatic development is that transcription factor NKX2.2 is detected in these progenitor cells of the mouse but not human (Jennings *et al.*, 2013).

As the multipotent pancreatic progenitor cells continue to proliferate there is a segregation and change in morphogenesis of the pancreatic buds (Shih, Wang and Sander, 2013). The buds elongate and as the gut tube rotates, the ventral and dorsal buds come into contact and fuse together. At CS19, multiple small protrusions then begin to form from the edge of the of the pancreatic bud (Gittes, 2009). The divergence into “tip” or “trunk” progenitor cells is marked by the expression levels of GATA4 (Jennings *et al.*, 2013). Tip cells that are located toward the edge of these protrusions are in the tip domain and express PTF1A, C-MYC and CPA (Shih, Wang and Sander, 2013). Tip domain cells are fated to become acinar cells. As development progresses acinar cells will continue to proliferate and increase in number by duplication (Pan and Wright, 2011). Trunk cells, which commit to central duct-like structures, express lower levels of GATA4 as compared to the more peripheral clustered pro-acinar tip cells, even though both progenitor cell types express several common pancreatic markers such as PDX1, SOX9 and NKX6.1 (Figure 1.2). Cells that are located toward the inside of the buds in the trunk domain are also characterized by the expression of NKX6.2, TCF2,

HNF6, PROX1 and HES1. Trunk domain cells are bipotential and give rise to endocrine and ductal cells (Shih, Wang and Sander, 2013).

The trunk domain consists of ductal and endocrine progenitor cells. The ductal progenitors form the primitive duct (or also known as the epithelial cord), which is characterized by tubules lined by a single layer of polarized epithelial cells. The commencement of endocrine specification is marked by the transient expression of transcription factor NGN3 (Figure 1.2). Cells that lack the expression of NGN3 will become the ductal cells. NGN3 positive endocrine precursors delaminate from the primitive ducts, converting to non-epithelial cells through a process thought to involve an epithelial-to-mesenchymal transition (Pan and Wright, 2011; Shih, Wang and Sander, 2013). In humans, NGN3 expression is detected at CS21 (8 wpc) around the end of the first trimester of human pregnancy upon the formation of fetal β-cells (Piper *et al.*, 2004; Lyttle *et al.*, 2008). These cells migrate into the surrounding area and coalesce into aggregates that become the islets of Langerhans (Gittes, 2009). At 10 to 14 wpc, NGN3 expression is at its peak level. This expression declines around 18wpc and is detected at extremely low levels after 35wpc (Salisbury *et al.*, 2014). SOX9 is absent in cells robustly expressing NGN3 and continues to be absent in subsequent endocrine cells, but is present in pancreatic duct cells (Jennings *et al.*, 2013). Each endocrine precursor cell will further differentiate into one of the five types of hormone expressing cells. At 12-13 wpc, islets containing α-cells, β-cells, δ-cells and γ-cells are apparent. α-cell fate is determined by the expression of ARX while β-cell identity is determined by PAX4, PDX1 and NKX6.1 (Shih, Wang and Sander, 2013).

1.3 Key transcription factors of human pancreas development

Transcription factors are gene regulatory proteins that play an integral role in governing gene regulatory networks and determining cell fate. During human pancreas development certain transcription factors play a key role in directing cell fates by regulating the transcription of genes involved in specification and ultimately mature function. Much of what is known of transcription factors in pancreas development has been revealed in rodent model systems such as genetically manipulated mouse models and cell lines (Conrad, Stein and Hunter, 2014). Although these models are great tools, there are notable distinctions between mouse and human pancreas, with implications in development and function. Some of the transcription factors that play a vital role in promoting human pancreatic development, function and identity are described below.

1.3.1 FOXA2

FOXA2 is a member of the winged helix/forkhead transcription factors and has been reported to be necessary for DE formation (Dufort *et al.*, 1998). FOXA2 transcription factor is consistently expressed from week 4 forward during development (Jennings *et al.*, 2013). FOXA2 acts as a pioneer factor to regulate PDX1 expression in mice. Recently this has been demonstrated in hPSC models of pancreas development as well (Lee *et al.*, 2002, 2019). FoxA2-/- mice have been demonstrated to be unable to develop the foregut and midgut tissues (Dufort *et al.*, 1998). FOXA2 persists in all mature pancreatic cell types of both mice and humans.

1.3.2 SOX17

SOX17 is a member of the HMG (high mobility group) box group of transcription factors. In humans, SOX17 is observed immediately before 4 weeks and is then excluded from pancreatic cells about 1 week later. This pattern of expression is similar to the down-regulation of *sox17* during mouse pancreatic development (Piper *et al.*, 2004). Mice models have indicated that early *sox17* expression is necessary for endoderm formation and it represses pancreatic fate at later stages (Spence *et al.*, 2009). During early development, SOX17 is regulated by the Wnt/ β-catenin signaling pathway, which promotes transcriptional activation of its target genes. SOX17 has also been shown to be a direct regulator of FOXA1 and FOXA2 (Sinner *et al.*, 2004). Sox17-/- mice and hPSCs do not form DE. Additionally, SOX17 is required for the induction of PDX1 expression during pancreas specification (Kanai-Azuma *et al.*, 2002).

1.3.3 HNF1B

HNF1B is a protein of the homeobox-containing basic helix-turn-helix family. HNF1B expression can be detected at 7 weeks in humans, and persists throughout pancreatic development (Jeon *et al.*, 2009). Heterozygous loss-of-function HNF1B mutations result in diabetes, termed MODY5, in humans. In mice only homozygous mutations show the same phenotype suggesting a higher sensitivity to HNF1B dosage levels in humans (Horikawa *et al.*, 1997). This could also be due to a potentiated single wave of human endocrine differentiation versus the two phases observed in mice. Hnf1b null mice die before gastrulation due to defective visceral endoderm formation. Using tetraploid

complementation, it has been shown that Hnf1b null mice displayed pancreas agenesis by e13.5 characterized by absence of the ventral pancreatic bud and an extremely reduced and transient dorsal bud (Haumaitre *et al.*, 2005).

1.3.4 PDX1

PDX1, also known as insulin promoter factor 1 (IPF1), has been studied for its role throughout all phases of pancreatic development. PDX1 is known to be critical during early pancreatic development and is expressed in all pancreatic precursor cells (Bernardo *et al.*, 2009). PDX1 is broadly expressed at around 4 weeks with a high level of expression being restricted later to adult human β cells (Lyttle *et al.*, 2008; Jennings *et al.*, 2013). In mice, pdx1 is expressed at e8.5 in the primitive gut tube and it marks the pre-pancreatic endoderm as well as parts of the stomach and duodenum. Pdx1 high expression then becomes restricted mostly to rodent β cells by e15.5, where it regulates the expression of *Ins1*, *MafA*, and *Pdx1* (Guz *et al.*, 1995; Offield *et al.*, 1996; Stoffers *et al.*, 1997).Low PDX1 expression is observed in the exocrine pancreas (Pan and Wright, 2011). In humans, PDX1 appears slightly later than in mice. Expression is only evident after the notochord and aorta are separate from the dorsal foregut in humans (Jennings *et al.*, 2013). Homozygous loss-of-function mutations in PDX1 result in pancreatic agenesis. Autosomal recessive mutations in the PDX1 locus have also been reported to cause permanent neonatal diabetes (Brissova, Blaha, *et al.*, 2005). PDX1 is commonly used as a marker for early pancreatic progenitors during the differentiation of hPSCs.

1.3.5 NKX6.1

NKX6.1, a member of the homeobox family of transcription factors, is a key regulator of pancreas development. The absence of nkx6.1 expression in deficient mice leads to impaired β-cell differentiation (Sander *et al.*, 2000). Expression of NKX6.1 is detected in the human pancreas from week 7 of development, gradually increasing from week 7 to week 21 (Jeon *et al.*, 2009). During human development, NKX6.1 is first detected, most readily in the dorsal pancreatic bud, at CS13 (30–33 dpc). By CS15 (35–37 dpc), NKX6.1 is co-expressed with PDX1, FOXA2 and SOX9 in pancreatic epithelial cells. This profile of NKX6.1, PDX1, SOX9, and FOXA2 distinguishes multipotent pancreatic progenitors in hPSC differentiation protocols (without NKX6.1, a combination of FOXA2, SOX9, and weak PDX1 detection could indicate extrahepatic biliary duct cells) (Jennings *et al.*, 2013). Additionally, in the adult human pancreas, the maintenance of β-cell identity is associated with the presence of PDX1 and NKX6.1 and changes in their expression and/or localization have been described in the islets of type 2 diabetic individuals. One theory suggests that the loss of β-cell identity possibly contributes to β-cell dedifferentiation (i.e. the regression to a progenitor-like state) in this disease (Marchetti *et al.*, 2017).

1.3.6 PTF1A

PTF1A is a basic helix-loop-helix (bHLH) transcription factor that forms part of the PTF complex. PTF1A plays a critical role in early pancreas specification from foregut endoderm (Kawaguchi *et al.*, 2002). In humans, PTF1A is detected at as early as 9 dpc

in the pancreas buds. From 14 dpc and on, it PTF1A is expressed only in the exocrine part of the pancreas. PTF1A was shown to play a key role in acinar development in mice, as homozygous ptf1a null mutant mice fail to develop acinar cells (Krapp *et al.*, 1998). Evidence from these early studies suggested that Ptf1a is a specific determinant for exocrine pancreatic cells only. However, lineage tracing studies in mice revealed that most pancreatic multipotent progenitor cells (MPCs) express Ptf1a, and are capable of differentiating into acinar, ductal or endocrine cells (Kawaguchi *et al.*, 2002; Pan *et al.*, 2013). When Ptf1a activity is totally inactivated, most of the MPCs are converted into non-pancreatic cell fates, such as gut and gallbladder. Haploinsufficiency or low levels of Ptf1a promote endocrine cell fates while repressing exocrine cell fates, and vice versa (Dong *et al.*, 2008). These findings underscore the distinct roles of Ptf1a in the specification and expansion of pancreatic progenitor cells. In later development and adulthood, Ptf1a promotes acinar differentiation and regulates acinar cell-specific gene expression, which is dependent on the PTF complex (Jin and Xiang, 2019). In humans, heterozygous mutations in *PTF1A* or a downstream enhancer of PTF1A cause pancreas agenesis (Sellick, Barker, Stolte-dijkstra, *et al.*, 2004; Weedon, Inês Cebola, *et al.*, 2014).

1.3.7 SOX9

SOX9 is a member of the SRY/HMG box family. In mice, Sox9 is highly expressed in both emerging pancreatic buds at e9.5 when they comprise almost exclusively of multipotent progenitors (Seymour, 2014). Sox9 also colocalizes with Pdx1 in early (by e9.0) dorsal and ventral pancreatic endoderm. By e12.5, Sox9 expression is barely

evident at the periphery, denoting Sox9 withdrawal from the distal tips. Concordantly, Sox9 becomes confined to trunk epithelium. In line with this proximal expression pattern, the distribution of Sox9 signal at birth is interpreted as marking a subset of ductal cells, islets, and a few acinar cells. The pattern of expression has led to Sox9 being considered a pancreatic progenitor marker (Seymour, 2014). In humans, it has been shown that SOX9 is weakly coexpressed with PDX1 in prospective duodenal-pancreatic endoderm by CS12 (29-31 dpc), equivalent to ~e9-9.5 in mouse embryogenesis (Jennings *et al.*, 2013). By CS13 (30-33 dpc ≈e9.5-10 in mouse), SOX9 is strongly expressed in the dorsal and ventral pancreatic buds and is maintained through CS16 (37-40 dpc ≈e12.25-12.75) in the branched pancreatic epithelium. The process of segregating SOX9 to the proximal trunk appears to occur relatively late during the human pancreatic program, between 10-14 wpc (weeks post-coitum). SOX9 rarely colocalizes with late endocrine progenitor markers and is excluded from all differentiated endocrine and acinar cells. This is consistent with SOX9 being enriched in pancreatic progenitors while being largely excluded from endocrine progenitors and differentiated endocrine and acinar cells (Seymour, 2014).

1.3.8 RORα

RORα is part of a family of nuclear receptors, also consisting of RORβ and RORγ, that can function as transcription factors to regulate gene expression (Chai *et al.*, 2013). They bind as monomers to the consensus DNA motif RGGTCA and can act as an “orphan” nuclear receptor to activate transcription even in the absence of bound ligands (Chauvet *et al.*, 2011). Ligand binding could lead to the recruitment of either co-activator

or co-repressor complexes, which suggests that RORα could act as either a repressor or activator of gene expression (Solt and Burris, 2012). A study of human embryos from late CS12 to early CS14 using laser capture of the developing pancreas and liver have suggested a specific role for RORα in regulating pancreas development (Jennings *et al.*, 2017). Of the 655 transcription factors identified as key pancreatic regulators in this study, 44% were predicted to be regulated by RORα using motif discovery. These studies highlight the potential role for RORα as a regulator of the pancreatic program in humans. Phenotypes observed in RORα mutant mice revealed a role for RORα in modulating diet-induced obesity, insulin sensitivity and glucose uptake (Lau *et al.*, 2011; Billon, Sitaula and Thomas P Burris, 2017). Studies have also shown the expression of ROR receptors in pancreatic islet cells and RORα regulates the expression of Ins2 in rat INS-1 cells (Mühlbauer *et al.*, 2013; Kuang *et al.*, 2014). Moreover, RORα was identified as a diabetes susceptibility locus in Mexican Americans and Han Chinese (Hayes *et al.*, 2007; Zhang *et al.*, 2016).

1.3.9 GATA6

GATA6 is a part of the GATA family of transcription factors that bind to the consensus sequence A/T-GATA-A/G. GATA4, GATA5 and GATA6 are the GATA family members found in endoderm derived tissues, the heart and gonads (Molkentin, 2000; Viger *et al.*, 2008). GATA6 consists of 7 exons with a total of 3770 base pairs; it contains two transactivation domains followed by 2 zinc finger DNA binding domains and a nuclear localization signal (Figure 1.3). GATA6 has two transcriptional start sites, both located in

the second exon resulting in the production of two isoforms of GATA6, one which is 595 amino acids in length and the other 449 amino acids long (Brewer *et al.*, 1999). The human and mouse GATA6/Gata6 gene was first described in 1996 (Morrisey *et al.*, 1996; Suzuki *et al.*, 1996). In humans, the GATA6 gene is located on human chromosome 18 q11.1–q11.2 (Suzuki *et al.*, 1996). During development, GATA6 is expressed in the primitive streak, heart, lung and pancreatic tissues (Koutsourakis *et al.*, 1999; Liu *et al.*, 2002; Sartori *et al.*, 2014). In adult tissues, GATA6 transcripts were expressed at high levels in the heart, ovary, lung, and pancreas, low levels in the liver and spleen. GATA4 expression often overlaps with GATA6 expression in the adult pancreas and heart (Suzuki *et al.*, 1996).

Figure 1.3: Structure of the GATA6 gene and protein

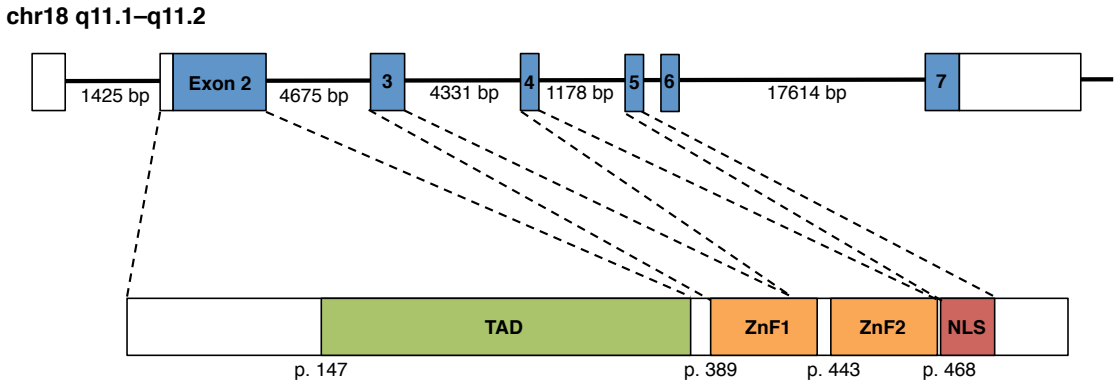


Figure 1.3: Structure of the GATA6 gene and protein
Representation of the gene structure of GATA6 (top) and protein (bottom). TAD, Transcriptional activation domain; ZnF, Zinc finger DNA binding domain; NLS, Nuclear localization signal.

GATA6 plays a critical role in human pancreatic development. While investigation of GATA6 in human development has been limited, it is expressed throughout the

pancreatic progenitor cell population at CS16-CS18 (Cebola *et al.*, 2015). During in-vitro differentiation, GATA6 is expressed from the primitive streak throughout the differentiation towards pancreas progenitors. During the later stages of development, GATA6 becomes restricted to the endocrine pancreas and ductal cells (Decker *et al.*, 2006). Loss of GATA6 in mice does not affect glucose tolerance or insulin levels when examined as adults (Carrasco *et al.*, 2012; Xuan *et al.*, 2012). Additionally, a conditional loss of GATA6 in adult β-cells leads to increased endoplasmic reticulum stress and minor β-cell death but does not have any effect on β-cell mass or glucose homeostasis (Sartori *et al.*, 2014). This suggests that GATA6 is not required in the endocrine compartment in mice during the later stages of development and in the pancreas of adult mice, GATA6 plays a non essential role in β-cell functionality. However, GATA6 is required in the exocrine pancreas for the maintenance and functionality of acinar cells. In mice with a lack of GATA6, the acinar cells are gradually lost resulting in the majority of the pancreas being replaced by fat (Martinelli *et al.*, 2013).

Recently, GATA6 has been revealed to be the most common cause of PA in a cohort of 27 neonatal diabetic patients with pancreatic agenesis or severe pancreas hypoplasia (Lango Allen *et al.*, 2012). Patients with a heterozygous mutation in GATA6 have been found with a range of phenotypes from no pancreatic defects to adult onset diabetes and pancreatic agenesis(De Franco, Shaw-Smith, Sarah E. Flanagan, *et al.*, 2013). While, a majority of GATA6 mutations in patients with pancreatic agenesis are de novo mutations, in rare cases where the mutation is inherited. However not all patients with the same mutation display the same phenotype suggesting incomplete penetrance. There have been studies of pancreatic agenesis patients and family members with

identical GATA6 mutations who have adult onset diabetes or even no abnormalities of the pancreas (Bonnefond *et al.*, 2012; De Franco, Shaw-Smith, Sarah E. Flanagan, *et al.*, 2013; Yu *et al.*, 2014).

Mice models have proven to be useful in understanding the role of GATA6 in development. However, mice with heterozygous loss of GATA6 are fertile and phenotypically normal (Morrisey *et al.*, 1998; Carrasco *et al.*, 2012; Xuan *et al.*, 2012). Mice with homozygous GATA6 null alleles have shown that GATA6 is essential for extra embryonic endoderm formation and leads to embryonic lethality (Morrisey *et al.*, 1998). However, it has also been shown that GATA6 null ES cells can form definitive endoderm and contribute to the primitive gut tube (Koutsourakis *et al.*, 1999; Zhao *et al.*, 2005). In mice, pancreas specific deletion of both GATA6 and GATA4 leads to pancreatic agenesis, suggesting a species-specific dosage requirement of these GATA factors (Carrasco *et al.*, 2012; Xuan *et al.*, 2012). In humans a heterozygous loss of GATA6 results in more severe phenotypes in comparison to murine models. Haploinsufficiency of GATA6 may be occurring in human patients while this may not be the case in murine models. Additionally, GATA4 and GATA6 may have greater functional redundancy in the mouse as compared to humans and thus in mouse models, a heterozygous loss of GATA6 can be compensated for with GATA4. To study a loss of GATA6 in a human based system we chose to use the differentiation of hPSCs.

1.4 Diseases of the human pancreas

There are various pancreatic diseases that affect humans. As our understanding of the human pancreas has increased, these diseases have been studied extensively and many therapies and potential cures have been developed. The major disease related to the endocrine pancreas is diabetes, where patients lose their sensitivity to glucose leading to higher blood glucose levels. There are many different forms of diabetes that will be discussed below. Additionally, some forms of pancreatic disorders are caused by developmental defects leading to neonatal diabetes mellitus (NMD). In some cases these developmental defects lead to patients being born with minimal to no pancreatic tissue. This is termed as pancreatic hypoplasia or pancreas agenesis (PA). The advent of hPSCs to model human pancreas development has vastly expanded our knowledge of developmental disorders associated with the pancreas.

1.4.1 Diabetes Mellitus

Diabetes is one of the oldest known diseases in human history. Indian texts dated from the 5th century BC describe people with symptoms of polyuria and increased thirst. In the past, diagnosis of diabetes was based on tasting the urine, as patients with this disorder would excrete glucose in the urine. In 1674, Thomas Willis coined the name diabetes mellitus, when he tasted dried urine from diabetic patients and described it as sticky and similar taste to honey (quasi melle) (Eknoyan and Nagy, 2005). Diabetes mellitus is rapidly becoming one of the largest and most expensive risks to global public health (Whiting *et al.*, 2011; Chan and Luk, 2016). Currently, approximately 8.5% of the

world population has some form of diabetes, and this prevalence is projected to continue to increase (Wareham and Herman, 2016). The global prevalence of diabetes has nearly doubled since 1980, rising from 4.7% to 8.5% in the adult population and causing 1.5 million deaths in 2012 (World Health Organization, 2016). It is estimated that 366 million people were diagnosed with diabetes in 2011; by 2030 this will have risen to 552 million (International Diabetes Federation, 2011). In 2017 alone, treatment costs related to diabetes in the United States rose to 327 billion dollars (Association, 2018). Unfortunately, diabetes has significant associated mortality, ranking as the sixth leading cause of death in the world and contributing to the mortality of other conditions such as cardiovascular and renal disease (Ashcroft and Rorsman, 2012; Zheng, Ley and Hu, 2018).

Diabetes mellitus results from impaired insulin secretion or action, or both, resulting in high serum glucose levels (Seino, Shibasaki and Minami, 2011). Despite decades of research, our understanding of the underlying pathogenesis of diabetes, especially pancreatic β-cell dysfunction, remains incomplete. In humans, physiological levels of blood glucose prior to meals are typically 3.9 mmol/L to 7.2 mmol/L (or 70 to 130 mg/dL) and two hours after a meal should not be higher than 10 mmol/L (or 180 mg/dL). Blood glucose levels in the body are regulated by the release of the antagonistic hormones from α-cells and β-cells. During regular homeostasis, when blood glucose is low, α-cells secrete glucagon stimulating the hepatic conversion of glycogen to glucose, which is then released into the bloodstream. When blood glucose levels are elevated, β-cells respond by releasing insulin causing the uptake, utilization and storage of glucose. Furthermore, insulin also subdues the hepatic production of glucose resulting in the

lowering of blood glucose levels (Saltiel and Kahn, 2001; Rhodes, 2005). Diabetes Mellitus can be categorized into four main categories, type I diabetes, type II diabetes, gestational diabetes and diabetes as a result of other causes. Other causes of diabetes include monogenic mutations, diseases of the exocrine pancreas and drug or chemical induced diabetes (American Diabetes Association, 2016).

1.4.1.1 Type One Diabetes Mellitus (T1DM)

T1DM, or juvenile-onset diabetes, accounts for approximately 5-10% of diabetic patients and is caused by the chronic autoimmune destruction of insulin-secreting β cells, usually leading to complete insulin deficiency or hyperglycemia. Hyperglycemia occurs when blood glucose levels are high, and this can lead to serious health conditions such as ketoacidosis, kidney failure, heart disease, stroke, and blindness. T1DM cases account for 10% of the diabetic population, and it is the most common form of diabetes in children <15 years of age. 90000 children are diagnosed every year, and currently, ~600000 children suffer from T1DM (Diaz-Valencia, Bougneres and Valleron, 2015).

The pathogenesis of T1DM and its molecular mechanism underlying this disease is not entirely clear. However, one of the characteristics of T1DM is the recognition of islet autoantigens by autoreactive CD4 (+) and CD8 (+) T cells and autoantibodies. Autoantibodies against islet autoantigens are involved in autoantigen processing and presentation by human leukocyte antigen (HLA) molecules. T cells that are specific for β-cell autoantigens are detectable in patients with advanced T1DM, and are able to recognize post-translationally modified peptides specific to β-cells (Katsarou *et al.*,

2017). The early phase of this disease is asymptomatic as the pancreas has an excess of β-cells. Autoantibodies can be detected at this stage, with disruption of glucose homeostasis being seen only with a significant loss of mass of β-cell mass (van Belle, Coppieters and von Herrath, 2011).

It is uncertain how T1DM is triggered, but after more than four decades of research, different aspects have been considered, from genetics to environmental. First, T1DM is a polygenic disease where genetic factors are necessary but not sufficient to cause the disease. There is a concordance rate of 30-50% in monozygotic twins and 6-10% in zygotic twins. An individual with a first degree relative with T1DM has a risk of 1 in 20, and in the general population that ratio is 1 in 300, indicating the influence that genetics plays while also demonstrating that even in a permissive genetic background, the disease is not 100% penetrant (Atkinson, 2012). There are around 50 loci associated with a susceptibility to develop the disease, with the HLA genes providing the most substantial contribution of approximately 60%. HLA genes encode for molecules that participle in antigen presentation and it is thought that HLA predisposes disease by presenting specific autoantigens. Many other loci are being investigated to gain new insights into genetic risk and autoimmunity (Concannon, Rich and Nepom, 2009).

Environmental factors are another component to may contribute to T1DM. One hypothesis suggests that stressors in the environment cause the β-cells to over-work, leading to exhaustion and β-cell damage, which trigger the immune response. Another theory suggests that the prevalence of T1DM is higher in countries where continuous improvement in sanitation and living standards lead to underused components of the

immune system attacking itself (Kondrashova *et al.*, 2013). Microbial infection with antigens cross-reactive to autoantigens may also yield the autoreactive T cells (Coppieters *et al.*, 2012). Another hypothesis called North-South Gradient is based on the observation that the highest incidence of T1DM is in Northern European countries, suggesting that low serum vitamin D could be associated with the development of the disease (Hypponen *et al.*, 2001). While many environmental factors have been proposed, the real cause of T1DM may be a combination of both genetic and one or more environmental factors.

1.4.1.2 Type Two Diabetes Mellitus (T2DM)

T2DM also results in disrupted glucose homeostasis caused by impaired insulin secretion, insulin resistance or both, as well as chronic inflammation (DeFronzo *et al.*, 2015). T2DM is characterized by insulin resistance, where insulin-sensitive target tissues such as the pancreatic β-cells, liver, muscle or adipose cells do not respond adequately to normal levels of insulin produced by intact β-cells. Consequently, this leads to disruption of the pancreatic β-cell function and decreased β-cell mass. Although T2DM is most prevalent in adults, there has been increasing evidence of this form of diabetes affecting younger individuals (Fajans, Bell and Polonsky, 2001). T2DM accounts for 90% of the reported diabetic cases, which are around 400 million adults that range from 20-70 years old (Zheng, Ley and Hu, 2018). Epidemiological studies show that development of T2DM is associated with sedentary behavior, lack of sleep or oversleeping and smoking cigarettes.

Genetics may also play a role in causing T2DM (DeFronzo *et al.*, 2015). Twin studies have shown that the heritability of T2DM could be 20-80% depending on the population examined. The risk of developing T2DM in a lifetime is 40% for individuals with one affected parent and this increases to 70% when both parents have it (Ali, 2013). Genome-wide association studies (GWAS) have also been used to identify risk genes associated with T2DM. A number of common variants have been identified in approximately 100 genes related with obesity (FTO, MC4R), cyclin-dependent kinase (CDKAL1, CDKN2A, CDKN2B), transcription factors associated with β-cell function and pancreas development (NEUROD1, HNF4A, HNF1A, TCF7L2, HHEX, PTF1A, PAX6, PAX4), ion channels (SLC30A8, SLC16A1) and others (Ali, 2013; Fuchsberger *et al.*, 2016). It has not only been challenging to identify the underlying genetic causes, but also to devise universal therapeutic strategies for T2DM. No cure has yet been found for the disease. Several forms of treatment that currently exist, especially for obese patients, include lifestyle modifications, treatment of obesity, oral hypoglycemic drugs, and insulin sensitizers such as metformin that reduces insulin resistance.

1.4.1.3 Monogenic Diabetes

Patients with monogenic diabetes have a mutation in a single gene involved in β-cell development or functionality. These mutations can be inherited as dominant or recessive, or arise spontaneously. Monogenic diabetes occurs approximately in 1 to 2 in 100,000 live births (Greeley *et al.*, 2010; Anik *et al.*, 2015), which accounts for 1%-4% of diabetes that are diagnosed in children (Rubio-Cabezas *et al.*, 2014). Monogenic diabetes can be subcategorized into neonatal diabetes and maturity onset diabetes of the young (MODY).

Neonatal diabetes mellitus (NDM) is a rare type of diabetes that occurs in the first six months of life with an incidence of 1 in 160,000 live births. Usually, the mutation occurs in genes related to pancreas development, insulin regulation or apoptosis. 8 different genes have been implicated in NDM (IPF1, EIF2AK3, GCK, FOXP3, KCNJ11, ABCC8, PTF1A and GLIS3) (Harries *et al.*, 2006; Ashcroft and Rorsman, 2012; Hattersley and Patel, 2017). The most common causes of NDM are activating heterozygous mutations in ABCC8 and KCNJ11. These genes encode for the K-ATP channel in β-cells. The unique sensitivity to ATP/ADP level changes induced by glycolysis makes this channel a metabolic cell sensor regulating insulin secretion (Hattersley and Patel, 2017). The K-ATP channel promotes insulin secretion. When the channel is open, K+ ions efflux to maintain the polarization of the cell membrane. When there is a change in ATP/ADP ratio in the cell, the K-ATP channel closes and the cell membrane is depolarized which triggers the influx of calcium and leads to a release of insulin granules. Activating heterozygous mutations in ABCC8 or KCNJ11 reduces the ability of the K-ATP channel to close and thus preventing glucose-induced electrical activity and insulin release which results in NDM (Ashcroft and Rorsman, 2012). Patients with neonatal diabetes can be divided into two subcategories: those that have transient neonatal diabetes mellitus which resolves after a few weeks to months and those who have permanent neonatal diabetes mellitus (Greeley *et al.*, 2010).

Maturity onset diabetes of the young (**MODY**) is the most common form of monogenic diabetes accounting for 1-2% of the total diabetes cases reported. MODY is characterized by heterozygous mutations in a single gene involved in the fetal

development of the pancreas or regulation of maturation and maintenance of beta cell function. MODY was first characterized in the 1970's and was described in patients with mild or no disease progression (Tattersall, 1974). Patients did not require insulin for treatment. In the 1990s the molecular characterization of different MODYs was revealed and since then a total of 13 types of MODY have been described in the literature. The four most common types of MODYs result from mutations in either HNF1α or MODY3 (30% to 50% of the cases), GCK or MODY2 (20% to 50% of the cases, HNF1β or MODY5 (5% of the cases) or HNF4α or MODY1 (5% of the cases) (Steck and Winter, 2011). Patients with MODYs typically do not have the presence of pancreatic or islet autoantibodies; have low insulin requirements and detectable c-peptide levels (Rubio-Cabezas *et al.*, 2014).

1.4.2 Pancreas Agenesis (PA)

Pancreatic agenesis (PA) is a rare disorder resulting from defective pancreatic development. The prevalence of PA is less than 1 in 1,000,000 and around 50 cases have been reported in the literature so far. Clinically, PA can be complete or partial, i.e. the pancreas can either be totally absent or extremely reduced in size (pancreatic hypoplasia). PA patients are diagnosed with NDM within the first 6 months of life and have severe pancreatic exocrine insufficiency. Patients also suffer from intrauterine growth retardation (IUGR) as a result of reduced insulin secretion in utero and are diagnosed with hyperglycemia in the first days of life. These patients require insulin treatment and exocrine enzyme replacement therapy. Diagnosis of pancreatic agenesis can be made by imaging (MRI or ultrasound) showing reduction or absence of

pancreatic tissue and measurement of fecal elastase that is often undetectable in patients with pancreatic agenesis as a result of exocrine dysfunction.

Isolated incidences of PA have been described as early as 1969 (Dourov and Buyl-Strouvens, 1969). The first genetic cause for PA to be identified was an inactivating mutation in *PDX1* (Stoffers *et al.*, 1997). The patient had a homozygous point mutation that caused a frameshift and showed signs of exocrine pancreas deficiency as well as hyperglycemia requiring insulin. Subsequently, other *PDX1* mutations have also been implicated in PA. The second gene noted to cause PA was pancreas specific transcription factor 1a (*PTF1A*). PTF1A plays a fundamental role in early pancreas specification from foregut endoderm (Kawaguchi *et al.*, 2002).

In humans, a mutation in *PTF1A* resulting in truncation of the C-terminal region was identified in a case of pancreatic and cerebellar agenesis (Sellick, Barker, Stolte-Dijkstra, *et al.*, 2004). Interestingly, a few cases of PA have recently been attributed to alterations in a region 25 kb downstream of *PTF1A* (Weedon, Ines Cebola, *et al.*, 2014; Gabbay *et al.*, 2017). This enhancer regulates the expression of PTF1A in human embryonic pancreatic progenitor cells, and mutations in the enhancer leads to its inactivation and PTF1A deprivation in these cells. One rare case of neonatal diabetes mellitus and pancreatic agenesis was also reported with compound heterozygous mutation of PTF1A (Gabbay *et al.*, 2017). The patient had a deletion and frameshift mutation in the PTF1A coding region on one chromosome, and a point mutation in the 400-bp enhancer region on another chromosome.

Figure 1.4: Pancreas Agenesis

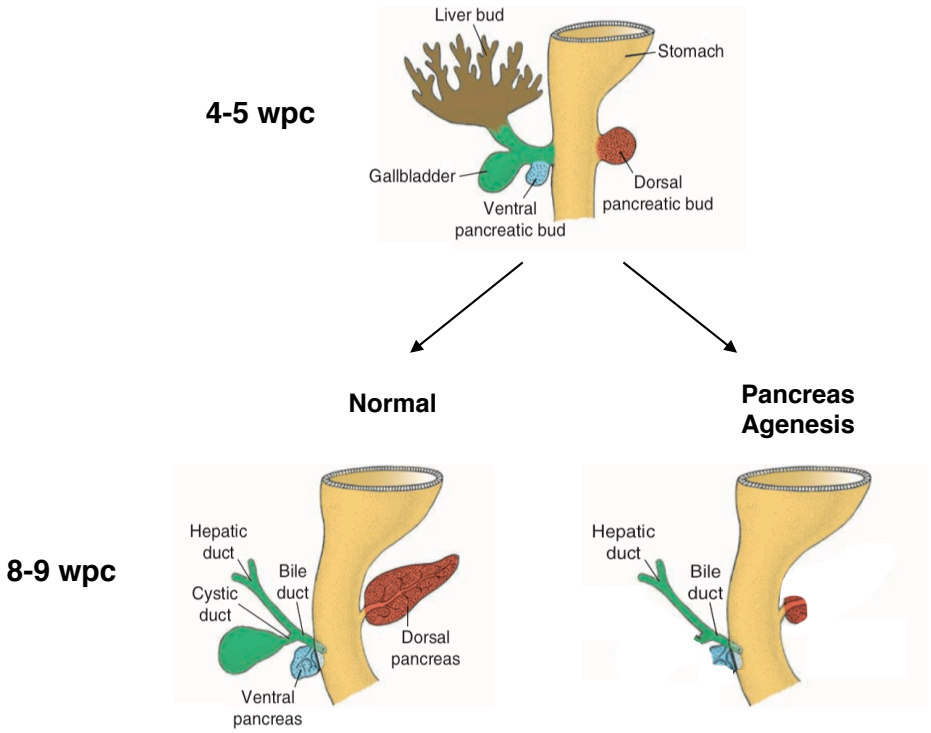


Figure 1.4: Pancreas Agenesis

During human development, the pancreatic buds appear around 4-5 weeks post coitum (wpc) and develop into the dorsal and ventral pancreas by 8-9 wpc. In patients suffering with PA, the dorsal and ventral panreatic regions are absent and patients are born with minimal or no pancreatic tissue. Adapted from Langman's medical embryology, 12th ed.

More recently, haploinsufficiency of *GATA6*, caused by heterozygous mutations, was shown to be the most common cause of PA (Lango Allen *et al.*, 2012). In these patients, pancreatic agenesis is commonly associated with other extrapancreatic malformations such as cardiac malformation, neurocognitive defects, hypothyroidism, gut abnormalities and gallbladder agenesis/biliary atresia (De Franco, Shaw-Smith, Sarah E Flanagan, *et al.*, 2013). Investigation of GATA6 in human development has been limited, with its

expression throughout the pancreatic progenitor cell population having been described at a single time period, CS16-CS18 (Cebola *et al.*, 2015). However, not all patients with GATA6-inactivating mutations causing congenital heart disease show PA and diabetes, and not all mutations arise *de novo* in the affected individual, implying some variability in the clinical phenotype (Bonnefond *et al.*, 2012). Conditional inactivation of *Gata6* and *Gata4* in mouse and the data imply a degree of redundancy between the two factors (Carrasco *et al.*, 2012; Xuan *et al.*, 2012). This redundancy between the GATA factors seems less apparent in humans. However, a few cases of *GATA4* mutation resulting in PNDM and at least some degree of exocrine insufficiency have also been described (Shaw-Smith *et al.*, 2014).

1.5 Pluripotent stem cells:

Pluripotent stem cells (PSC) are cells that have the ability to indefinitely self renew in an undifferentiated state and have the potential to be differentiated into any cell type in the body. Embryonic stem (ES) cells are cells that are derived from the inner cell mass of the blastocyst while induced pluripotent stem (iPS) cells are generated by introduction of pluripotency factors. PSCs have the potential to give rise to unlimited supplies of functional human cells that can be used in the study of developmental biology, disease mechanisms, drug discovery and therapeutics (Girlovanu *et al.*, 2015). The first described isolation of human ES cells from the blastocyst was in 1998. The human ES cells were derived from the inner cell mass of the human blastocyst and could be maintained as colonies on a feeder layer (Thomson *et al.*, 1998). The long-term maintenance of these cells has been studied for many years and the optimal culture

conditions for the maintenance of pluripotency with basic fibroblast growth factor (bFGF) are well established, allowing the growth and propagation of human ES cells in in vitro cultures (Yamanaka, 2012). Human ES cell lines were shown to have high telomerase activity, be karyotypically stable and displayed pluripotency as they are able to form all three germ layers in a teratoma assay. Human ES cells express the transcription factors NANOG, OCT4 and SOX2. They also express the extracellular markers SSEA-3, SSEA4, TRA-1-60 and TRA-1-81 (Thomson *et al.*, 1998).

Given the relative limited access to ES cells from human embryo blastocysts, a lot of effort has been put into inducing pluripotency in somatic cells, which culminated in the development of induced pluripotent stem (iPS) cell technology. In 1962, the first reprogramming of cells by somatic cell nuclear transfer was demonstrated. A nucleus from the intestinal epithelium cell of a swimming tadpole stage was transferred to an unfertilized egg and shown that the egg could successfully develop into a tadpole demonstrating that development could be reverted to an earlier stage (Gurdon, 1962). Additionally, in 1989, the transcription factor myosin D was identified as a master regular whose expression could convert the cell fate of multiple cell lines to muscle cells (Weintraub *et al.*, 1989). This was a very early demonstration that the introduction and expression of a transcription factor could change cell fate. Based upon these ideas, in 2006, the generation of iPS cells from both embryonic and adult mouse fibroblast cells was first described. 24 transcription factors that are known to play roles in maintaining pluripotency we screened by retroviral expression and this list was narrowed down to 4 necessary factors; oct3/4, sox2, c-myc and klf4 (Takahashi and Yamanaka, 2006). Their work demonstrated that adult differentiated cells could be reprogrammed back to an

embryonic stem cell state. The following year, human iPS cells were derived from adult human fibroblasts using the same four transcription factors that were used to generate mouse iPS cells (Takahashi *et al.*, 2007).

The discovery of iPS cells was a large advancement in the field, as it provides a potential for studying pluripotency without the ethical complications that may be associated with the derivation of human ES cells and also a method to generate patient specific stem cells. hPSCs have the potential to be used in disease models for studying the molecular basis of diseases, including genetically inherited human diseases (Yamanaka, 2012). It brings tremendous potential not only in disease modeling, but also in regenerative medicine, cell replacement therapy, drug testing and targeted gene-repair strategies, such as homologous recombination to repair genetic defects. Thus, they serve as ideal model systems for human developmental scientific studies.

1.6 Modeling pancreas development using stem cells

Using the knowledge derived from studying the embryonic development of various model organisms, detailed protocols have been established to derive pancreatic cells from hPSCs. The in vitro differentiation of hPSCs is achieved by sequential exposure to growth factors based on known signaling pathways and mechanisms of development (Figure 1.5). During in vivo embryonic development the process of gastrulation leads to the formation of the primary germ layers, ectoderm, mesoderm, and endoderm (Tam and Behringer, 1997). In the mouse embryo, the primitive streak forms during

gastrulation in the posterior side of the embryo at the border between the epiblast and extraembyronic tissue. Uncommitted epiblast cells that will eventually become the mesoderm and definitive endoderm (DE) germ layer undergo an epithelial to mesenchymal transition and migrate through this structure. Epiblast cells that migrate through the more posterior region of the primitive streak will form the mesoderm while cells that migrate through the more anterior region of the primitive streak will form DE. In vitro primitive streak differentiation is first established by the addition of the canonical Wnt ligand, Wnt3a, or small molecules that can mimic this signal. Cells are then differentiated to DE by the activation of Nodal signaling, a ligand for transforming growth factor-β (TGF-β) superfamily, through the addition of high concentrations of the Nodal surrogate, Activin A (Kubo *et al.*, 2004; Gadue *et al.*, 2006). Activin A is another TGF-β family member and induces a similar intracellular signal as Nodal. The first stages of the pancreatic differentiation protocols involve the differentiation of primitive streak cells into DE cells using elevated concentration of Activin A (D'Amour *et al.*, 2005).

As development of the endodermal germ layer proceeds, the primitive gut tube forms organs buds that will proliferate and eventually differentiate into the mature organs. In 2006, a comprehensive stepwise pancreatic specification protocol was introduced, describing the differentiation of hESCs to endocrine cells with the use of specific growth factors and chemical compounds (D'Amour *et al.*, 2006). With this in vitro differentiation protocol, the cells mimic in vivo pancreas organogenesis by being directed through stages resembling DE, gut-tube endoderm, pancreatic endoderm and endocrine precursor, thus recapitulating the major stages of normal pancreatic endocrine development. Each stage is recognized by the expression of specific markers. Soon

after, improved pancreas differentiation protocol using serum-free media, retinoic acid (RA) and nicotinamide was published (Jiang *et al.*, 2007). The protocol had 4 distinct stages: DE formation, pancreatic endoderm formation, endocrine induction and islet-like cluster formation. These cells were produced insulin but showed limited responsiveness to glucose stimulation.

This study was followed by various modified variations of pancreatic differentiation protocols (Kroon *et al.*, 2008; Zhang *et al.*, 2009; Loh *et al.*, 2014; Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; M Cristina Nostro *et al.*, 2015; Russ *et al.*, 2015). The Melton group described the generation of glucose-responsive cells from hPSCs following an approach of coaxing cells through different stages mimicking pancreas development. Cells were coaxed to NKX6–1+/PDX1+ cells by employing previously published protocols. From NKX6–1+/PDX1+ cells, β-cell-like cells (termed SC-β cells, ~33% NKX6–1+/C-peptide+) emerged in 2–3 weeks. Non-NKX6–1+/C-peptide+ cells in the final population were α-cells, δ-cells and PDX1+ pancreatic progenitors that had not differentiated into endocrine cells. SC-β cells responded to 2–3 successive glucose challenges in vitro by releasing insulin with a comparable stimulation index (ratio of insulin secretion at 20 mM to 2 mM glucose) to that of primary human adult islets. Similar to the aforementioned report, Kieffer and colleagues devised a 7-stage protocol for driving the fate of hPSCs to pancreatic endocrine cells. Cells differentiating to pancreatic endoderm progeny expressing PDX1/NKX6–1 (stages 1–4) were grown in a planar culture and then they were plated as clusters onto filter inserts (air-liquid interface (ALI) culture) for the remaining stages. More recently, refinements of the protocol, most notably the omission of TGFβ signaling inhibitors at the β-cell maturation stage, now permit the generation of

glucose-responsive β-cells in vitro (Nair *et al.*, 2019; Velazco-Cruz *et al.*, 2019; Veres *et al.*, 2019). It is important to note that these refined protocols were optimized to predominantly produce β-cells for use in cell therapy and that other pancreatic cell types are not generated to the same extent as seen during embryonic development. This could limit the utility of these protocols for studying developmental mechanisms.

Despite significant recent advances in the differentiation of hPSCs into pancreatic cells, several challenges remain to obtain functional human pancreatic – particularly insulin-producing – cells performing on par with native cells. The variability noted in different reports in the specification outcomes across various hPSC lines points to the need for protocol adaptation and optimization, generally hindering direct translation to multiple lines. In addition, glucose-stimulated insulin secretion (GSIS) of many stem cell-derived pancreatic cells differs significantly from that of primary human islets. More importantly, the underlying causes of these discrepancies between native islet cells and hPSC-derived cells are unclear. The ability to differentiate hPSCs to functional β-cells in vitro is important for the study of the mechanisms of β-cell failure. PSCs with specific mutations can be obtained through genome editing of ES cells or from diabetic patient derived iPS cells. In vitro derived β-cells can also be utilized to investigate β-cell proliferation, maintenance, in drug testing and provide the potential for the generation of an unlimited supply of β-cells if these cells are to be used in tissue replacement therapy.

Figure 1.5: Differentiation of hPSCs to β-like cells

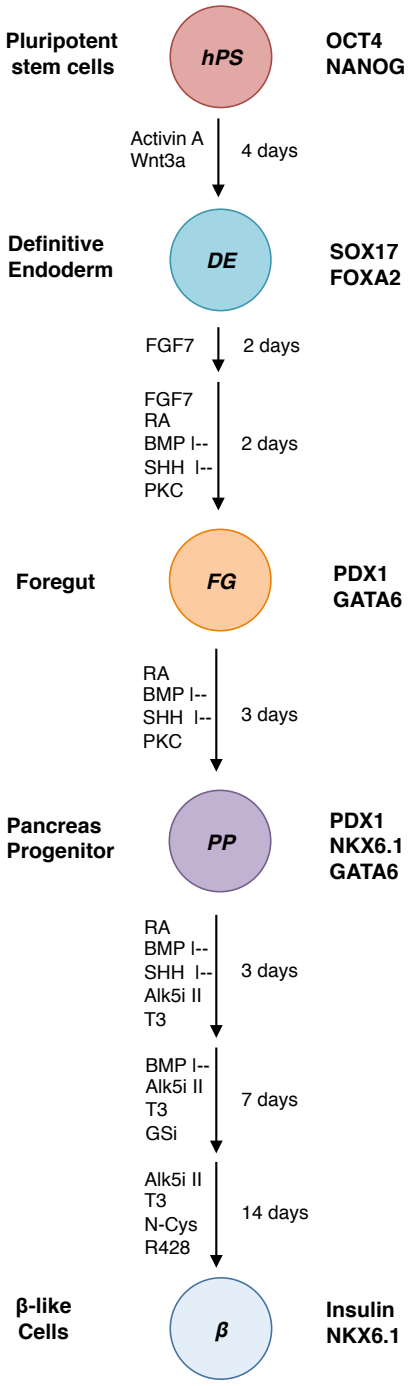


Figure 1.5: Differentiation of hPSCs to β-like cells

An illustration showing the different stages, signaling pathways and timeline of human pancreas differentiations from hPSCs. Key transcription factors that are used to identify cells at each stage are indicated.

1.7 Genetic modifiers of disease penetrance

While many of the strongest associations between genotype and phenotype are coding mutations or variants, in a significant number of diseases there exists inter-individual differences, termed 'variable penetrance' (Castel *et al.*, 2018). Variable penetrance and variable expression of genes are common phenomena that often cause individuals carrying the same variant to display highly variable symptoms. This is true even in the case of Mendelian and other severe diseases driven by rare variants that have a strong effect on phenotype. These phenomena are a key challenge for understanding how genetic variants manifest in human traits and a major practical caveat for the prognosis of an individual's disease outcomes based on their genetic data. However, the causes and mechanisms of variable penetrance are poorly understood. In addition to environmental modifiers of genetic effects, a potential cause of variable penetrance involves other genetic variants with additive or epistatic modifier effects (Cooper *et al.*, 2013). While some studies have successfully mapped genetic modifiers of, for example, BRCA variants in breast cancer (Milne and Antoniou, 2011) and RET variants in Hirschsprung's disease (Emison *et al.*, 2005), the cause of variable penetrance in a large number of diseases are unknown. In part, this is because large-scale genome-wide association studies (GWAS) typically lack power and are easily affected by confounders (Wei, Hemani and Haley, 2014). A targeted analysis of a specific variant or gene that is strongly implicated in a rare disease typically suffers from a low number of carriers and lack of an ideal model system for testing these diseases modifiers. A large-scale study in 2018 showed that non-coding regulatory variants could affect the penetrance of pathogenic coding variants. By using diverse data types from population and disease

cohorts, the study provided strong evidence of modified penetrance due to joint functional effects of regulatory and coding variants. Importantly, this suggested that the combination of an individual's regulatory and coding variant genotypes has an effect on phenotype, since purifying selection acts only on traits that affect fitness (Castel *et al.*, 2018).

Figure 1.6: The effects of non-coding regulatory variants on disease penetrance

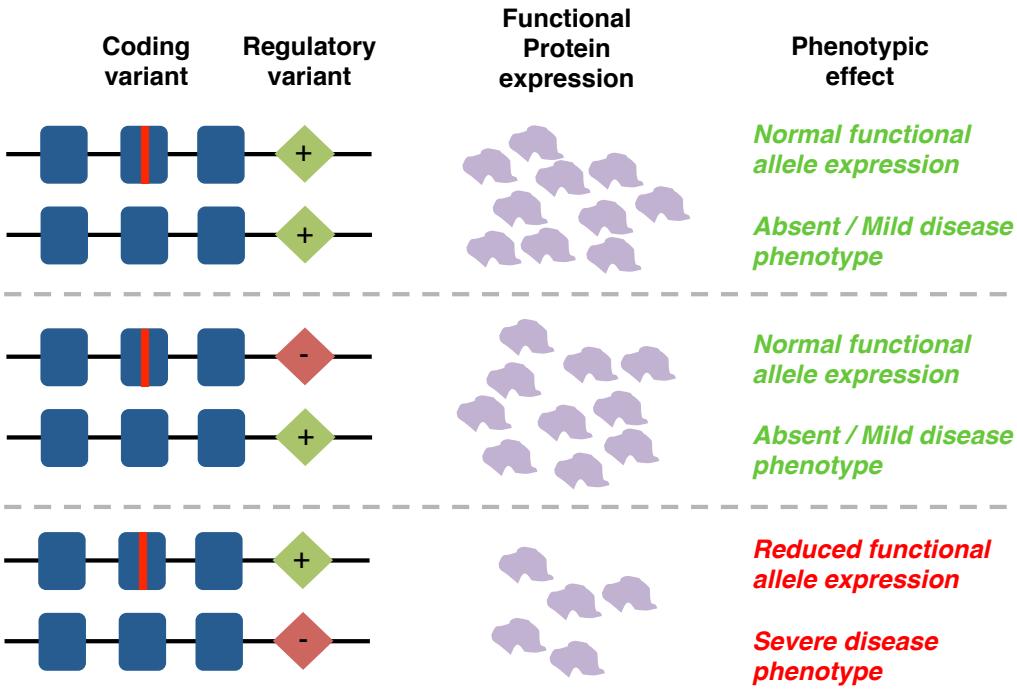


Figure 1.6: The effects of non-coding regulatory variants on disease penetrance

In certain monogenic diseases that display a variability in penetrance, the presence of non-coding regulatory variant could affect functional protein expression and manifestation of the disease. In the above illustration, if the non-coding variant is present in *cis* with the non-functional coding allele, the functional allele produces enough functional protein to prevent a severe disease phenotype. However, if the non-coding variant is present in *trans* with the non-functional coding allele, the amount of functional protein produced from both alleles is reduced, resulting in a severe disease phenotype.

Interestingly, most of the single-nucleotide variants/polymorphisms (SNVs/SNPs) that provide statistical evidence for increased risk of complex diseases have been mapped to non-coding regions (Zhang and Lupski, 2015). While the risk variants revealed by GWAS may be linked with a neighboring causative coding variant, an alternative and more likely explanation is that functionally relevant genetic variants reside in non-coding regions with functional consequences on nearby genes. Accumulating evidence has shown that the GWAS variants are significantly enriched in the functional non-coding regions such as enhancer elements, DNase hypersensitivity regions and chromatin marks (Ahonen *et al.*, 2009; Degner *et al.*, 2012; Trynka *et al.*, 2013). Functional non-coding variants can affect *cis*- or *trans*-regulatory elements, suggesting potential genome interaction that can be investigated by various genome technologies. Chromosome Conformation Capture (3C) enables the study of chromatin looping and genome architecture in three dimensions; 3C has been used for functional annotation of the schizophrenia-associated loci suggested by a genome-wide association study (Ripke *et al.*, 2013; Roussos *et al.*, 2014). The schizophrenia-associated SNPs identified by GWAS are located in the predicted enhancers downstream of *CACNA1C*. They map to where the 3C assay in human dorsolateral prefrontal cortex and iPSC-derived neurons revealed promoter–enhancer interaction through chromosome loops.

While computational methods enhanced by the growth of GWAS have been an invaluable tool, functional assays are important experimental steps to investigate the molecular mechanism of non-coding variants in human disease. A classic approach to study the functional consequences of human disease-associated SNPs has been to engineer variation into mouse models.

In the past few years, the renovation and development of the CRISPR/Cas technology enables the quick and accurate genome editing in various model organisms for both SNPs and small indels (Shalem, Sanjana and Zhang, 2015). While evolutionary conservation is still the prerequisite for functional assays of non-coding variants in model organisms, when the non-coding variants of interest are not conserved in mouse, it has also been feasible to directly obtain the human genetic variant of interest using the stem cell technologies, such as patient-derived iPSCs, for further *in vitro* functional assays (Donnelly *et al.*, 2013; Sareen *et al.*, 2013).

1.8 Studying genetic modifiers of disease penetrance using stem cells

Since their discovery over 10 years ago, iPSCs have been used to model a multitude of "diseases in a dish" by utilizing lines derived from a relatively small number of diseased and healthy donors (Avior, Sagi and Benvenisty, 2016; Warren *et al.*, 2017). iPSCs can be directly derived from the somatic cells of any donor prospectively identified to carry a genotype of interest (Park *et al.*, 2008). This versatile feature makes iPSCs an ideal platform for the study of Mendelian or monogenic diseases. The differentiation of human iPSCs to cell types of relevance offers a powerful platform with enormous potential for disease modeling, drug screening, and cell therapeutics. The combination of human iPSC technology with genome editing technique allows the generation of isogenic cell lines that differ in single genetic changes for causal modeling of candidate variants and genes, offering a new tool linking genotypes to phenotypes in the study of human cell biology and disease (Fang *et al.*, 2019).

A good example of using iPSCs to model disease modifying SNPs is from the study by Soldner et al. that demonstrated functional connect of GWAS-identified risk variants of Parkinson's disease in neurons derived from human iPSCs (Soldner *et al.*, 2016). They focused on Parkinson's disease associated risk SNPs, which were located in a α-synuclein (SNCA) regulatory region based on genome-wide epigenetic information. They were able to identify subtle changes in allele-specific transcription of SNCA between two SNPs located in the SCNA enhancer region. As a follow up approach, they used CRISPS/Cas9 to knock-out a single allele of the SNP and determine how the SNP affected SNCA expression. They found that allele-specific expression roughly translated to an increase of total SNCA expression in neurons and in neural precursors. Furthermore, sequence-dependent binding of the brain-specific transcription factors EMX2 and NKX6-1 on this locus was revealed.

Another example of disease modeling using iPSCs is presented by Ebert et al. (Ebert *et al.*, 2014). They studied a SNP in the gene coding for aldehydronease 2 enzyme, which confers a loss of cardioprotective effects and increases the risk for coronary artery and ischemic heart disease. Cardiomyocytes (CM) differentiated from iPSCs derived from an east-Asian population genotyped for a common ALDH2* SNP (MAF = 0.08), demonstrated that CMs carrying the ALDH2* genotype had increased levels of oxidative stress and aldehyde byproduct 4HNE buildup. Accumulation of these two byproducts resulted in dysregulated cell cycle and apoptosis signaling, which exacerbated damage and reduced cellular recovery to ischemic challenge in the CMs of ALDH2* carriers,

thereby establishing the cellular mechanisms for increased disease susceptibility for a single SNP.

Hypertension (HTN), a common, complex and polyfactorial disease, that can be affected by genetic variants which regulate blood pressure was studied by Hamazaki et al. (Hamazaki *et al.*, 2017). Animal models have been used intensively for studying systemic diseases like HTN, however they may not always be suitable for understanding the biological impact of human genetic variants. It is also difficult to obtain a large number of appropriate tissues of relevance for the phenotype of interest (e.g., vascular smooth muscle or endothelium) from a person with a specific genotype to test the biological or functional consequences of these genetic variations. To combat such challenges, Biel et al. constructed an iPSC repository from 17 HTN patients, whose genome-wide SNP variations as well as clinical responses to antihypertensive drugs were available (Biel *et al.*, 2015). They then differentiated these iPSCs into vascular smooth muscle cells and quantified their contraction in response to various physiological stimuli. Furthermore, the study also demonstrated the ability of iPSCs to recapitulate a SNP-associated modification of PRKCA expression.

It is increasingly important to understand how specific risk variants functionally contribute to underlying pathogenesis. Compared with single gene mutation found in monogenic diseases, the effects of SNP variants can often be minor or subtle. It is important to utilize isogenic cells to decode the significance of such gene variants.

Recent advances in genome-editing technology (e.g., CRISPR/Cas9 systems) have simplified the ability to target specific genetic loci for functional studies.

1.9 Overview of Research Goals

While recent studies have implicated GATA6 as the most common cause of pancreatic agenesis, the molecular mechanisms that regulate the development of this disease are still not entirely understood. Additionally, the variability in penetrance of GATA6 heterozygous mutations raises questions regarding other factors that may influence GATA6 levels that could lead to variability in phenotypes. The goal of this project is to answer these questions and develop a deeper understanding of pancreatic agenesis caused by GATA6 haploinsufficiency. The finding that a patient derived iPS line had a more severe developmental phenotype as compared to an ES cell line with the same GATA6 mutation piqued our interest in this project. Given the majority of known *GATA6* mutations have been identified by sequencing the coding region of the gene, we hypothesized that this variable disease penetrance is caused by differences in the non-coding region of the *GATA6* locus that regulate its expression during development (Rodriguez-Segui, Akerman and Ferrer, 2012; Yorifuji *et al.*, 2012; Catli *et al.*, 2013; Suzuki *et al.*, 2014; Yu *et al.*, 2014; Christina S Chao *et al.*, 2015; Stanescu *et al.*, 2015). We identified a non-coding SNP rs12953985, which lies approximately 8kb downstream of the *GATA6* gene, with the patient line harboring the minor allele variant (A) of this SNP. Interestingly, we also found that the minor allele variant of this SNP was enriched in a cohort of 33 total patients with pancreatic agenesis caused by *GATA6* mutations. Using CRISPR-CAS9 genome editing, we analyzed the effect of this variant on GATA6 expression during *in vitro* pancreas differentiation from multiple genetically matched

hPSC lines. We found that the minor allele variant of rs12953985, in conjunction with a *GATA6* heterozygous mutation, led to the largest reduction in GATA6 protein expression specifically during pancreas specification and a more severe defect in generating pancreatic progenitors. We also determined that the minor allele variant of this SNP impaired RORα binding by chromatin immune-precipitation (ChIP) assay. Inhibiting the function of RORα using an inverse agonist decreased GATA6 expression and inhibited pancreas development. This effect was observed only in lines with an intact RORα binding site supplied by the major allele variant of the SNP. Finally, we modified the minor allele variant of the SNP to a consensus RORα binding site in the *GATA6* PA patient iPSC line. We found that the iPSC line where both the coding mutation was corrected and the RORα binding site is introduced has the highest levels of GATA6 expression and most efficient pancreas differentiation capacity. This study identifies a possible genetic modifier contributing to the pancreatic agenesis phenotype in patients with *GATA6* mutations.

CHAPTER 2: MATERIALS AND METHODS

2.1 EXPERIMENTAL MODEL AND SUBJECT DETAILS

2.1.1 hPSC lines

The Mel1 ES cells were obtained from Ed Stanley and Andrew Elafanty at the Murdoch Children's Research Institute(Micallef *et al.*, 2012). The CHOP.Panagenesis1 (patient IPS+/indel) IPSC cells were generated from a lymphoblastoid cell line by reprogramming using episomal vectors by the Stem Cell core at the Children's Hospital of Philadelphia. The CHOPWT6 iPSC line was also reprogrammed by the Stem cell core at the Children's Hospital of Philadelphia and was previously published as the WTBM1-8 iPSC line (Sullivan *et al.*, 2014).

2.1.2 Monogenic Diabetes Registry

Patients Chicago#1-5 and N6320 were consented for the Monogenic Diabetes Registry, an IRB-approved longitudinal study of people with monogenic diabetes housed at the University of Chicago. Over 3,000 participants are currently enrolled in the Registry and other associated studies. Information about the Monogenic Diabetes Registry can be found at www.monogenicdiabetes.org.

2.1.3 *GATA6* testing at the Exeter Molecular Genetics laboratory

The probands of Exeter families 1 to 31 had been referred to Exeter for genetic testing. The *GATA6* gene (NM_005257.4) was tested either by PCR followed by Sanger sequencing as previously described (De Franco *et al.*, 2013) or by targeted next generation sequencing of all the known monogenic diabetes genes (Ellard *et al.*, 2013). Clinical information was provided by the referring clinicians via a neonatal diabetes request form (available at www.diabetesgenes.org) or from clinical notes. The study was conducted in accordance with the Declaration of Helsinki principles with informed parental consent given on behalf of children.

2.2 METHOD DETAILS

2.2.1 hPSC culture

hPSC cell lines were cultured on 0.1% gelatin and irradiated mouse embryonic fibroblast (MEF) feeder cells in DMEM/F12 supplemented with 2mM of glutamine, 15% Knockout Serum Replacement (KSR), 1X NEAA, penicillin/streptomycin, 0.1mM β-mercaptoethanol and 10ng/ml of bFGF. The medium was changed every day. Cells were passaged when they reach 80% confluence, approximately every 4 days, using TrypLE at a 1:6 ratio. In all hPSC cultures, 5 μM Rho-associated protein kinase (ROCK)

inhibitor Y-27632 (Selleck Chemicals, #S1049) was only added into the culture media for ~18 hours when passaging or thawing hPSCs.

2.2.2 Generation of genetically modified hPSCs mediated by CRISPR-CAS9

The 19bp gRNA's of interest were cloned into the vector (Addgene, #41824) using In-Fusion HD (Clontech Cat No. 639647). PSCs were plated onto gelatin coated 6-well plates with MEF's 24 hours prior to transfection and were transfected at 40% to 60% confluency. 0.5µg CAS9-GFP (Addgene, plasmid #44719), 0.5µg gRNA, 0.5 µg of ssODN and 3µL Lipofectamine Stem reagent was diluted separately in 50µL DMEM-F12 and gently mixed together. The mixture was incubated at room temperature for 10 minutes and added dropwise into 1 well of cells. 18 to 24 hours post transfection, cells were harvested with TrypLE and cell sorted for GFP positive cells. These cells were plated at low density (~1000-2000 cells / 10cm dish) in human ES cell maintenance media with 5 µM ROCK inhibitor (Cayman chemicals) onto a 1:3 matrigel coated 10cm tissue culture dishes containing MEFs. Approximately 14 to 20 days later single colonies were manually picked and screened. To screen for positive clones, genomic DNA was extracted from the clones by resuspending the cells in 20 µL of Accutaq PCR buffer (Sigma-Aldrich) with 0.1 µg/mL proteinase K (Qiagen) and incubated at 55°C for 60 minutes followed by 95°C for 10 minutes. Cell debris was spun down and 5 µL of supernatant was used for PCR. The PCR products were analyzed by 3.5% agarose gel electrophoresis and sequencing. Heterozygous mutants were confirmed by sub-cloning PCR products using the TOPO TA cloning kit (Life Technologies) sub-cloning of cells to exclude potential contamination of cells with different genotypes (e.g., mixing of

homozygous mutant cells with WT cells). All primers, guide RNA's and single stranded DNA oligos used for genome editing are listed in Table 2-1.

Table 2-1. Guide RNA's, ssODN's and primers

Guide RNA's	**Sequence 5' to 3'**
gRNA for introducing 4bp duplication in exon 2	AGTGGGCCAGCCAACCACGCGGG
gRNA for introducing minor allele variant of rs12953985	GATGGAAGAGACTTGACTGAGG
gRNA for introducing RORα consensus	CTTGACTGAGGTCCAGCTGAGG
ssODN's	**Sequence 5' to 3'**
ssODN for introducing 4bp duplication in exon 2	TCCCCGGTCTACGTGCCCACCACCCGCGTGGGTTCCA TGCTGCCCGGCCTACCGTACCACCTGCAGGGGTCGGG CAGTCAGTGGGCCCGCGAATCATGCAGGCGGCGCGG GCGCGCACCCCGGCTGGCCTCAGGCCTCGGCCGACA GCCCTCCATACGGCAGCGGAGGCGGCGCGGC
ssODN wild-type exon 2	TCCCCGGTCTACGTGCCCACCACCCGCGTGGGTTCCA TGCTGCCCGGCCTACCGTACCACCTGCAGGGGTCGGG CAGTGGGCCCGCGAATCATGCAGGCGGCGCGGGCGC GCACCCCGGCTGGCCTCAGGCCTCGGCCGACAGCCCT CCATACGGCAGCGGAGGCGGCGCGGC
ssODN for introducing minor allele variant of rs12953985	TACTGACGAACCACAACCCAACCTTTCTGAGACTCAGT TTCCTCAGCT GGACTTCAGTCAAGTCTCTTCCATCTCTAAAGTTATATG AATTGATGA
ssODN for	TACTGACGAACCACAACCCAACCTTTCTGAGACTCAGT

introducing RORα consensus	TTCATCAGCTTGACCTCAGTCAAGTCTCTTCCATCTCTA AAGTTATATGAATTGATGAGCTGGACCTCAGTCAA
Primers	**Sequence 5' to 3'**
PCR primer's for 4bp duplication in exon 2	Forward- CGGAGGAGATGTACCAGACC Reverse- CTGGGAGAGTAGGGGAAGC
Sequencing primer's for 4bp duplication in exon 2	Forward- TTCGTGCACTCTGCG Reverse- TGAGGCCAGCCGGGGTGC
PCR primer's for R5	Forward- GGTGTTGTCTGATTAAGGAAATCTAGGTACC Reverse- ATAAATTATAAGTAACAAGCAACCGGTGTTCCC
Sequencing primer's for R5	Forward- GGAAATCTAGGTACCTCACACTGC Reverse- AAAACTTTGCCGAGTGTATCTGTC

2.2.3 Pancreatic differentiation from hPSC's

In all differentiation assays, mutants were analyzed in parallel with isogenic WT controls. hPSC's were passaged onto 1:30 matrigel coated 6 well plates using TryplE and ROCK inhibitor. Cells were fed with hPSC media every day until they reached ~90% confluency. Pancreas differentiation was initiated on day 0 with RPMI media supplemented with 3μM Chir99021 and 100μg/ml Activin A. On day 1 media was changed to RPMI with 100μg/ml Activin A, 0.3μM Chir99021 and 5μg/ml bFGF. Day 2 was SFD with 100μg/ml Activin A. Cells were harvested on day 3 (DE stage) for flow

cytometry analysis and RNA collection using 0.25% Trypsin for 5 minutes. From days 3 to 5 cells were fed with DMEM-F12 containing 0.25mM ascorbic acid, 50ng/ml FGF7 and 1.25μM IWP2. Day 6-8 media contained DMEM high glucose (5g/L) supplemented with 1:100 B27 without RA, 1X glutamax, 0.25mM ascorbic acid, 1:200 ITS-X, 50ng/ml FGF7, 0.5μM SANT-1, 1μM Retinoic Acid, 100nM LDN-193189 and 500nM Phorbol. For experiments with SR1001, the RORα inverse agonist, 1μM of SR1001 was added to the media from day 6-8. Cells were harvested on day 8 (PFG stage) for flow cytometry analysis and RNA collection. Media for days 9-11 consisted of DMEM high glucose (5g/L) supplemented with 1:100 B27 without RA, 1X glutamax, 0.25mM ascorbic acid, 1:200 ITS-X, 2ng/ml FGF7, 0.5μM SANT-1, 0.1μM Retinoic Acid, 200nM LDN-193189 and 250nM Phorbol. Cells were harvested on day 11 (PP stage) for flow cytometry analysis and RNA collection. From days 12-14 the media was changed to MCDB131 supplemented with 20mM glucose, 2% FBS, 1X Glutamax, 1:200 ITS-X, 10ug/ml Heparin, 10uM Zinc sulfate, 0.5μM SANT-1, 0.05μM Retinoic Acid, 200nM LDN-193189, 1μM T3 and 10μM ALK5i II. From day 14-28 cells were fed every other day with media that contained MCDB131 with 20mM glucose, 2% FBS, 1X Glutamax, 1:200 ITS-X, 10ug/ml Heparin, 10uM Zinc sulfate, 200nM LDN-193189, 1μM T3, 10μM ALK5i II and 100nM GSIS XX. From day 29-40 cells were fed every other day after with media that contained MCDB131 with 20mM glucose, 2% FBS, 1X Glutamax, 1:200 ITS-X, 10μg/ml Heparin, 10uM Zinc sulfate, 1μM T3, 10μM ALK5i II, 1mM N-acetyl cysteine, 10μM Trolox and 2μM R428. Cells were harvested on day 40 (β-like stage) for flow cytometry analysis and RNA collection.

2.2.4 Flow Cytometry

Single cell suspensions were prepared by treating cells with 0.25% Trypsin/EDTA for 3 to 5 minutes. For intracellular staining, cells were fixed with 1.6% paraformaldehyde (Electron Microscopy Science) for 30 minutes at 37°C. Cells were washed, permeabilized and stained with 1X saponin buffer (Biolegend). Primary antibodies were diluted to the appropriate concentrations in 100uL of saponin buffer and cells were stained for 30 minutes at room temperature. Samples were washed using 100uL saponin twice and incubated for 30min using the appropriated secondary antibody. Following the staining, cells were resuspended in FACS buffer (DPBS with 0.1% BSA and 0.1% sodium azide). All samples were run on a FACSCantos II or Cytoflex flow cytometer (Becton Dickinson) and analyzed using FlowJo (Treestar) software program.

2.2.5 RNA isolation and cDNA synthesis

Cells were lysed using Lysis buffer provided with the PureLink RNA Micro Kit (Invitrogen Cat No 12183-016) and stored at -80c. To harvest RNA, samples were thawed out at 4c and RNA was extracted using the PureLink RNA Micro Kit following the manufacturer's instructions. 14µl of RNAse free water was used to resuspend the isolated RNA. cDNA was produced using the SuperScript™ III First-Strand Synthesis System kit (Invitrogen). Quantitative PCR was carried out on a LightCycler 480 II with SYBR select master mix (Invitrogen). For all experiments, TBP (Veazey and Golding, 2011) was used as a

housekeeping gene to determine relative gene expression levels. Gene expression levels were then divided by wild type levels for better graphical representation. All primers used for qRTPCR are in Table 2-2.

Table 2-2. qRTPCR primers

Gene	Sequence 5' to 3'
CDX2	Forward- TCCTGGTCTGGGAAGGGAAGAGAAA Reverse- CGGAAGCCAAAGGCAGCTAAGATAG
GATA4	Forward- GCTGTGCTGTGGTGGGTTAAGT Reverse- CGCCCTGCATCCCTAATACCAAATC
GATA6	Forward- GAGGCTTGCTGAAAGAGTGAGAGAAGA Reverse- TCCTAGTCCTGGCTTCTGGAAGTG
NKX6.1	Forward- AGGACGACGACTACAATAAGCCTCTG Reverse- CGCTGCTGGACTTGTGCTTCT
PDX1	Forward- GCCGGCTCTTCAAAGACAATGGA Reverse- GGTCGCCCGAGTAAGAATGGCTTTAT
RORα	Forward- TACCTGGACATACAGCCTTC Reverse- CGTTGGTGAACGAACAGTAG
SOX2	Forward- ATGACCAGCTCGCAGACCTACA Reverse- GGACTTGACCACCGAACCCA
TBP	Forward- TTGCTGAGAAGAGTGTGCTGGAGATG Reverse- CGTAAGGTGGCAGGCTGTTGTT
R5-rs12953985 ChIP-qPCR	Forward- ACTTTAGAGATGGAAGAGACTTGAC Reverse- CTAGCCCACAGCACCAAG

PDX1 ChIP-qPCR	Forward- TATCCAGTCAGAGGCTGGTC Reverse- CTGACTTAAACAGGGCCACTT
FOXN1 ChIP-qPCR	Forward- CAGGCAGGATAAAGGTTGGA Reverse- TCTGGGACTTATGTGGAGAGG
SPARC ChIP-qPCR	Forward- TTGGGCCTGGTTCTGCCCCT Reverse- CCGGGGCTGCTGCCTAAACC
GATA4 ChIP-qPCR	Forward- CGCTGAGAGCACAGACAAT Reverse- AGGGAGTGGGAAAGACCA

2.2.6 Immunofluorescence staining

At days 2, 7 and 10 of the differentiation, cells were harvested by incubating with 0.25% Trypsin for 5 mins at 37c. These cells were spun down at 1200rpm for 3 mins in PBS and plated onto 1:3 matrigel coated glass coverslips in the appropriate differentiation media with 5 μM ROCK inhibitor. Cells were fixed the next day at DE (day 3), PFG (day 8) or PP (day 11) stages. For the NKX6.1 and SOX2 stains cells were fixed in 4% PFA in PBS for 15 minutes at room temperature. For the RORα staining, cells were fixed at -20c for 20 mins using cold methanol. Fixed cells were then washed 5-8 times with PBS on ice. Fixed cells were blocked for one hour (5% normal goat serum, 0.3% TritonX-100 in DPBS) and stained in primary antibody in staining buffer (1% BSA, 0.3% TritonX-100 in DPBS) overnight at 4°C. After washing with 3 times PBS for 5 mins each, cells were stained in secondary antibody in staining buffer for 2 hours at room temperature. After washing with 3 times PBS for 5 mins each, cells were stained with Hoescth diluted in

PBS for 15 mins. Slides were viewed under a Leica DMI 4000B microscope and digital images were captured with Leica Application Suite software.

2.2.7 Enhancer cloning and luciferase reporter assays

The 3' regulatory region, R5, was PCR amplified from genomic DNA of the $Mel^{+/+ \mid G/G}$ cell line with Phusion High-Fidelity DNA Polymerase (New England BioLabs) (primers in TableS2). These PCR products were cleaned up using NucleoSpin® Gel and PCR Clean-up kit (Machery-nagel) and cloned into a pGL4.23[*luc2*/minP] vector backbone (Promega) using In-FUSION HD. Correct cloning was assessed by Sanger sequencing and restriction enzyme digestion. DNA was prepared with the HiSpeed Plasmid Maxi Kit (Qiagen). At days 2 and 7 of the differentiation, cells were harvested by incubating with 0.25% Trypsin for 5 mins at 37c. These cells were spun down at 1200rpm for 3 mins in PBS and plated onto 1:30 matrigel coated 12 well plates in the appropriate differentiation media with 5 μM ROCK inhibitor. These cells were then transfected on the same day with either 1ug of pGL4.23-R5 vector or empty pGL4.23 vector and 4ng of *Renilla* normalizer control pGL4.75[*hRluc*/CMV] (Promega) using Lipofectamine Stem reagent (Invitrogen) in Opti-MEM (Gibco) according to the manufacturer's' instructions. Luciferase activity was measured 24 h after transfection (at the DE and PFG stages) with the Dual-Luciferase Reporter Assay System (Promega). Firefly luciferase activity was normalized to *Renilla* luciferase activity and then to the amount of empty pGL4.23[*luc2*/minP] vector backbone.

2.2.8 ChIP-qPCR

PP cells ($1x10^7$ cells) were harvested using 0.25% Trypsin and cross-linked in 1% formaldehyde in PBS by shaking for 10 min at room temperature. The cross-linking reaction was stopped by the addition of glycine to a final concentration of 125 mM and shaking for 5 min at room temperature. Cross-linked cells were washed 3X with ice cold PBS and pelleted by spinning at 2000rpm for 5 min at 4c. For chromatin fragmentation, cells were resuspended in 1ml cell lysis buffer (10 mM Tris-HCl (pH 8.0), 10 mM NaCl, 0.2% NP-40) with fresh protease inhibitor and PMSF for 10 min on ice. Cells were spun down and resuspended in 1ml nuclei lysis buffer (50mM Tris-HCL (pH 8.1), 10mM EDTA, 1%SDS) with fresh protease inhibitor and PMSF and sonicated in a Covaris S220 sonicator with a duty cycle of 2%, a peak incident power of 105 W and 200 cycles per burst for 20 min. The fragmented chromatin was diluted 1:2 in IP dilution buffer (20 mM Tris-HCl (pH 8.1), 2 mM EDTA, 150mM NaCl, 1% Triton X-100, 0.01% SDS, Protease Inhibitors) and directly used for immunoprecipitation. Samples were precleared with 5ug IgG isotype, Protein G agarose beads for 2 hours at 4c. 180ul of the supernatant was saved as Input DNA. The rest of the supernatant were split equally into 2 1ml tubes (5 X 10^6 cell each) for incubation with 10ug RORα or IgG control antibody. These antibodies were pre bound to Protein G agarose beads overnight at 4c. Beads were then washed 5 times using (1.) IP wash 1(20mM Tris-HCL (Ph 8.1), 2mM EDTA, 50mM NaCl, 1% Triton X-100, 0.1%SDS), (2 and 3.) High salt buffer (20mM Tris-HCL (Ph 8.1), 2mM EDTA, 500mM NaCl, 1% Triton X-100, 0.01%SDS), (3.) IP wash 2 (10mM Tris-HCL (Ph 8.1), 1mM EDTA, 0.25M LiCl, 1% NP-40, 1% deoxycholic acid), and (4 and 5.) TE buffer (10 mM Tris-HCl (pH 8.0), 1 mM EDTA). Subsequently, protein-DNA complexes were eluted

from the beads in Elution-Buffer (100 mM NaHCO3 and 1% SDS) at 65c for 20 min. Cross-links were reversed at 65c overnight and incubated with ProteinaseK and RNase for 2 hours at 55c. DNA was extracted using 400 µl of phenol/chloroform/isoamyl alcohol by vortexing then centrifuged at 14,000 rpm for 5 min at room temperature. The aqueous layer containing pulled down genomic DNA was transferred to fresh 1.5 ml Eppendorf tubes. 16uL of 5M NaCl, 1.5uL of 20mg/mL glycogen and 800µl of 100% ethanol were added to the samples, which were then vortexed. The samples were next incubated overnight at -20°C to precipitate the DNA. Precipitated DNA was pelleted by centrifuging at 14,000 rpm for 30 min at 4°C. The DNA pellet was then washed with ice-cold 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C. The DNA was resuspended with 100ul TE buffer and 1.5ul was used for each qPCR reaction.

In order to identify positive control regions for the ChIP, we looked for RORα binding motifs within active regulatory regions of known pancreas specific transcription factors, such as GATA4, PDX1 and SOX9. We also tested gene expression of SPARC and NR1D2 at the PFG stage. RORα has been demonstrated to bind to the regulatory regions of SPARC and NR1D2 in HepG2 cells (Chauvet *et al.*, 2011). We detected RORα binding at the PFG stage in a previously described regulatory region of SPARC as well as at a region proximal to GATA4 (Figure 3-9 D).

2.2.9 Chipseq data analysis

Raw read fastq files for PDX1 ChIPseq (GSE58686 Teo et. al.), H3K4Me1, Input, HNF6, FOXA2 and PDX1 ChIPseq (E-MATB-1990 Weedon et. al.) were downloaded. ATAC-seq raw read were downloaded from GSE114101 (Lee *et al.*, 2019). Next, reads were aligned to hg19 genome using bowtie2 (2.2.6) with only 1 mismatch allowed per read and only 1 alignment per read. These sam files were converted to bam format using samtools1.6. Next the bam files were converted to bedgraph format using bedtools2.27.1. We used MACS2 for peak calling with a p value cut-off of 1e-5. These peaks were verified by performing motif analysis using the findMotifsGenome.pl command on HOMER. Bedgraph files for Input, H3K4Me1 and peaks for PDX1, HNF6 and FOXA2 were loaded onto Intergrative Genomic Viewer (IGV v2.3) for visualization.

2.2.10 Western Blot

PFG and PP cells ($1x10^7$ cells) were harvested using 0.25% Trypsin, washed twice with PBS, pelleted and stored at -80c. Protein from cell pellet was quantified using Pierce™ BCA Protein Assay Kit (Thermo Fischer scientific, cat No. 23227). 20ug of protein was loaded onto a 4-12% Bis-Tris SDS-polyacrylamide gel (Invitrogen). Samples were transferred into a PVDF membrane (Thermo Fisher) and membrane blocking was performed using 2% nonfat dry milk. The membrane was stained in primary antibody diluted in 2% nonfat dry milk overnight at 4°C. After washing 3 times with 1X PBS-T for 5 mins each, the membrane was stained in secondary antibody diluted in 2% nonfat dry

milk for 1 hour at room temperature. The membrane was washed 3 times with 1X PBS-T for 5 mins each. HRP was detected using Pierce™ TMB Substrate Kit and membrane was exposed to HyBlot CL autoradiography film (Denville Scientific) to visualize the protein band.

2.2.11 SiRNA knockdown of RORα

Human RORα DsiRNAs (hs.Ri.RORA.13.2) and Scrambled negative control DsiRNA were obtained from Integrated DNA Technologies (IDT) . At day6 of the differentiation, cells were harvested by incubating with 0.25% Trypsin for 5 mins at 37c. These cells were spun down at 1200rpm for 3 mins in PBS and plated onto 1:30 matrigel coated 12 well plates at a 1:1 ratio in the appropriate differentiation media with 5 μM ROCK inhibitor. 10nM of the control scrambled siRNA or RORα siRNA (13.2) were added onto these cells at the same time with Lipofectamine RNAi MAX (Invitrogen) following the recommended procedure. SiRNAs were removed the next day by replacing with fresh media. Transfected cells were harvested and examined for knockdown efficiency 48 hrs after transfection (PFG stage) for RNA level by qRT-PCR and protein level by immunofluorescence. Cells were also fixed for intracellular flow cytometry as described.

2.3 QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical analysis was performed using the GraphPad Prism software. The results are expressed as the mean ± standard error of the mean. Ordinary one-way ANOVA with multiple comparisons were used for statistics with correction for multiple comparisons using statistical hypothesis testing perform using Tukey. An unpaired two-tailed Student's t-test with the assumption of same SD was performed for the enhancer luciferase assays. An unpaired two-tailed Student's t-test performed for the quantifying c-peptide+ cells at the β-like stage. In figures * $P<0.05$, ** $P<0.01$, *** $P<0.001$ and **** $P<0.0001$. Statistical analysis on the patient cohort data was performed using R. Chi-square and Fisher Exact probability tests were used to compare the frequency of the minor allele of the SNP in different groups. Odds ratios were calculated to estimate the risk of the minor allele A in the PA versus non PA groups.

Table 2-3. Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit anti-GATA6	Cell Signaling Technology	5851S
Biotinylated goat anti PDX-1/IPF1	R&D Systems	BAF2419
Mouse IgG1 anti-NKX6.1	DSHB	F55A10
Rat anti-Somatostatin	Santa Cruz	sc-47706
Mouse IgG1 anti-Glucagon	Sigma-Aldrich	G2654-.2ML
Rabbit anti-SOX2	Cell Signaling Technology	3579
Mouse IgG1 anti-SOX2	Biolegend	656102
Mouse IgG1 anti-SOX17	BD Pharmingen	561590
Rabbit anti-C-Peptide	Cell Signaling	4593S
Rabbit anti-CDX2	Abcam	ab76541
Hoechst 33342 solution	Thermo Fischer Scientific	62249
Rabbit anti-ROR alpha	Abcam	ab60134
Goat anti-mouse IgG1-488	Jackson Immunoresearch	115-545-205
Goat anti-mouse IgG1-PE	Jackson Immunoresearch	115-115-205
Goat anti-mouse IgG1-647	Jackson Immunoresearch	115-605-205
Goat anti-rabbit alexa 647	Invitrogen	A21245
Goat anti-rabbit IgG-PE	Jackson Immunoresearch	111-116-144
Donkey anti-mouse IgG alexa647	Jackson Immunoresearch	715-605-150

Donkey anti-rabbit IgG-PE	Jackson Immunoresearch	711-116-152
Streptavidin, Pacific Blue conjugate	Thermo Fischer Scientific	S11222
Goat anti Rat alexa 647	Thermo Fischer scientific	A21247
Chemicals, Peptides, and Recombinant Proteins		
Iscove's DMEM	Corning	10-016-CV
DMEM/F12	Corning	10-092-CV
2-β mercapthoethanol – 55mM	Invitrogen	21985023
NEAA – 10 mM	Invitrogen	11140050
Recombinant human bFGF	R & D Systems	233-FB/CF
Y-27632 (ROCK inhibitor)	R&D systems	1254/50
Fetal Bovine Serum	Tissue Culture Biologicals	101
Knock-out Serum Replacement	Invitrogen	10828-028
TRYPLE Express w/ Phenol Red	Invitrogen	12605010
Pen/Strep 100X	Mediatech	MT30-002-CI
L-Glutamine	Mediatech	MT25-005-CI
Gelatin	Sigma	G1890
Activin A	R&D	
CHIR99021	Tocris	4423
bFGF	Thermo Fischer Scientific	PHG0263
FGF7	R&D	251-KG-050/CF
SANT-1	Sigma-Aldrich	S4572
Retinoic Acid	Sigma-Aldrich	R2625
LDN-193189	STEMGENT	04-0074

Phorbol 12-myristate 13-acetate	Tocris	1201
Heparin	Sigma-Aldrich	H3149
Zinc Sulfate		
T3 (250 mg)	Sigma-Aldrich	T6397
ALK5 inhibitor	Enzo Life Sciences	ALX-270-445
GSIXX	Calbiochem	565789
N-Acetyl-L-cysteine	Sigma-Aldrich	A9165
Trolox	EMD	648471
SR1001	Tocris	4868
R428	Selleck Chemicals	S2841
Critical Commercial Assays		
Dual-Luciferase® Reporter Assay System	Promega	E1910
Invitrogen PureLink RNA Micro Kit	Invitrogen	12183-016
Gibson Assembly® Master Mix	NEB	E2611S
NucleoSpin® Gel and PCR Clean-up	Machery-nagel	#740609.250
HiSpeed Plasmid Maxi Kit	Qiagen	12663
SuperScript™ III First-Strand Synthesis System	Invitrogen	18080051
In-Fusion HD cloning kit	Clonetech	639647
Experimental Models: Cell Lines		
Provided in Table1		
Oligonucleotides		
Guide RNA's, ssODN's, primers and siRNA's	This thesis	Table 2-1
Primers for quantitative RT-PCR	This thesis	Table 2-2

Recombinant DNA		
pGL4.23[*luc2*/minP]	Promega	E8411
pGL4.75[*hRluc*/CMV]	Promega	E6931
pCas9_GFP	Addgene	44719
gRNA empty vector	Addgene	41824
Software and Algorithms		
FlowJo	Ashland	https://www.flowjo.com/solutions/flowjo/ downloads
GraphPad Prism	GraphPad Software	https://www.graphpad.com/support/faqid/%201952
ApE plasmid Editor	M. Wayne Davis	http://biologylabs.utah.edsen/wayned/ape/
Leica Application Suite X	Leica	https://www.leica-stems.com/products/microsftware/details/product/leica-
Bowtie2		http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Samtools 1.6		http://samtools.sourceforge.net/
Bedtools 2.27.1		https://bedtools.readthedocs.io/en/latest/
MACS2		http://liulab.dfci.harvard.edu/MACS/
HOMER		http://homer.ucsd.edu/homer/
Intergrative Genomic Viewer (IGV v2.3)		http://software.broadinstitute.org/software/igv/
SnapGene		https://www.snapgene.com/

CHAPTER 3:

THE PENETRANCE OF PANCREATIC AGENESIS CAUSED BY GATA6 MUTATIONS IS MODIFIED BY A NON-CODING SNP

3.1 Introduction

Pancreatic agenesis (PA) is a rare disorder resulting from defective pancreatic development. Clinically, PA can be complete or partial, but distinguishing between these two possibilities can be difficult. For the purposes of this work, we are defining PA as patients presenting with severe exocrine insufficiency and neonatal diabetes diagnosed in the first 6 months of life. Isolated incidences of PA have been described as early as 1969, however the genetic causes of this disease were not unraveled until 1997, when a homozygous mutation in *PDX1* was identified as a cause (Dourov and Buyl-Strouvens, 1969; Stoffers *et al.*, 1997). Since then, *PTF1A*, another key pancreatic transcription factor has been implicated as a cause of PA (Sellick, Barker, Stolte-dijkstra, *et al.*, 2004; Concepcion *et al.*, 2014; Shaw-Smith *et al.*, 2014). More recently, haploinsufficiency of *GATA6*, caused by heterozygous mutations, was shown to be the most common cause of PA (Lango Allen *et al.*, 2012). Interestingly, as is common with many disorders caused by haploinsufficiency, patients with *GATA6* heterozygous mutations display a large phenotypic variability, with the pancreatic phenotype ranging from PA, to adult-onset diabetes, to absence of diabetes even in adulthood (Shaw-Smith *et al.*, 2014; Shi *et al.*, 2017). In some cases, marked phenotypic variability has been observed even among affected members of the same family (Yau *et al.*, 2017). This phenotypic variability can be stochastic, due to mosaicism in the case of *de novo* mutations, and/or resulting from secondary disease-modifying genes that act in the GATA6 pathway during development

(De Franco, Shaw-Smith, Sarah E Flanagan, *et al.*, 2013; Shi *et al.*, 2017; Tiyaboonchai *et al.*, 2017). Given the majority of known *GATA6* mutations have been identified by sequencing the coding region of the gene, another possible explanation for disease penetrance could be differences in the non-coding region of the *GATA6* locus that regulate its expression during development (Rodríguez-Seguí, Akerman and Ferrer, 2012; Yorifuji *et al.*, 2012; Catli *et al.*, 2013; Suzuki *et al.*, 2014; Yu *et al.*, 2014; Christina S. Chao *et al.*, 2015; Stanescu *et al.*, 2015).

The increasing importance of non-coding regions in development and disease have been a fairly recent phenomenon (Zhang and Lupski, 2015). A rather surprising feature of the fully sequenced and annotated human genome has been the unexpectedly high proportion of functional regions of the genome. While only about 1.3% of the genome is ascribed to coding regions, close to 40% of the human genome has been speculated to have regulatory functions on tissue-specific gene expression (Heintzman *et al.*, 2007; Thurman *et al.*, 2012; Gordon and Lyonnet, 2014). Thus studying non-coding variants residing in these regulatory regions may play a pivotal role in understanding the incomplete penetrance of haploinsufficiency related disorders (Mcclellan and King, 2010; Brewer *et al.*, 2014; Castel *et al.*, 2018). Single nucleotide polymorphisms (SNPs) can alter transcription factor binding at regulatory regions and hence have a major impact on chromatin architecture and gene expression (Do *et al.*, 2017; Chen *et al.*, 2018). A number of such SNPs have been identified by genome-wide association studies (GWAS) to be associated with complex diseases (Visscher *et al.*, 2017). However mechanistic insights are limited due to the lack of adequate model systems to understand how specific risk variants functionally contribute to the underlying

pathogenesis (Mcclellan and King, 2010; Soldner *et al.*, 2016). The advent of clustered regularly interspaced short palindromic repeats (CRISPR)-CRISPR-associated 9 (Cas9) gene editing has proven to be an invaluable tool to test the effect of disease modifying SNPs on gene expression in specific tissue associated cell lines and diseases (Hsu, Lander and Zhang, 2014; Chen *et al.*, 2018; Wang *et al.*, 2018). However, studies testing the effect of SNPs on gene expression during embryonic development have been limited (Soldner *et al.*, 2016; Pashos *et al.*, 2017).

Human pluripotent stem cells (hPSCs) present a great opportunity to study developmental genetic disorders. The ability to have clonal populations and isogenic cell lines, where only a single mutation or SNP differentiates two cell lines, makes them ideal for teasing out the function of a given genomic change on development and disease (Musunuru, 2013). As our fundamental knowledge of developmental pathways expand, the increasing number of *in vitro* differentiation protocols to generate various cell types from hPSCs has grown considerably. Furthermore, in diseases such as PA, mouse models do not completely phenocopy the human disease. In fact, the loss of three or more alleles of *Gata6* and *Gata4*, the two major GATA family members expressed during pancreas specification, are needed for mice to mimic the human phenotype (Koutsourakis *et al.*, 1999; Zhao *et al.*, 2005; Carrasco *et al.*, 2012; Xuan *et al.*, 2012). Using directed differentiation protocols to model pancreas development from hPSCs, it has been shown that *GATA6* haploinsufficiency leads to a defect in generating pancreas progenitors and fewer pancreatic endocrine cells (Shi *et al.*, 2017; Tiyaboonchai *et al.*, 2017).

In this study, we investigated the role of the non-coding SNP rs12953985, which lies approximately 8kb downstream of the *GATA6* gene. We found that the minor allele variant (A) was enriched in a cohort of 33 total patients with pancreatic agenesis caused by *GATA6* mutations. Using CRISPR-CAS9 genome editing, we analyze the effect of this variant on GATA6 expression during *in vitro* pancreas differentiation from multiple genetically matched hPSC lines. We find that the minor allele variant of rs12953985, in conjunction with a *GATA6* heterozygous mutation, leads to the largest reduction in GATA6 protein expression specifically during pancreas specification and a more severe defect in generating pancreatic progenitors. We also determine that the minor allele variant of this SNP impairs retinoic acid receptor-related orphan receptor α (RORα) binding by chromatin immune-precipitation (ChIP) assay. Inhibiting the function of RORα using an inverse agonist decreases GATA6 expression and inhibits pancreas development. This effect is observed only in lines with an intact RORα binding site supplied by the major allele variant of the SNP. Finally, we modify the minor allele variant of the SNP to a consensus RORα binding site in the *GATA6* PA patient induced pluripotent stem cell (iPSC) line. We find that the iPSC line where both the coding mutation was corrected and the RORα binding site is introduced has the highest levels of GATA6 expression and most efficient pancreas differentiation capacity. Our study identifies a possible genetic modifier contributing to the pancreatic agenesis phenotype in patients with *GATA6* mutations and provides insights into RORα as a regulator of pancreas development in humans.

3.2 Results

3.2.1 GATA6 haploinsufficiency leads to a pancreas progenitor defect and a switch in cell fate

In recent years, *in vitro* pancreas differentiation protocols to generate β-like cells from hPSCs have become well established (Pagliuca *et al.*, 2014; Rezania *et al.*, 2014; M. Cristina Nostro *et al.*, 2015). These protocols leverage existing knowledge of developmental pathways to guide hPSCs through defined stages of development such as the SOX17+ definitive endoderm (DE), PDX1+ posterior foregut (PFG) and PDX1+/NKX6.1+ pancreas progenitor (PP) stages (Figure 3-1 A). Previous studies have shown that pancreas progenitors marked by PDX1 and NKX6.1 give rise to endocrine, ductal and acinar lineages in humans (Jennings *et al.*, 2013, 2015b). Given that a large proportion of patients with GATA6 heterozygous mutations are born with minimal to no pancreatic tissue, a defect in generating pancreas progenitors is expected. Recently, studies using genome edited embryonic stem cell (ESC) lines have shown that heterozygous mutations in GATA6 lead to a defect at the pancreas progenitor (PP) stage (Shi *et al.*, 2017). However, given the variability in the types and locations of mutations in *GATA6*, the use of iPSC lines derived from PA patients can aid in understanding the effects of different mutations. We used a previously described PA patient iPSC line (denoted as $iPS^{+/mut}$) to verify whether that particular *GATA6* mutation leads to the same phenotype at the PP stage (Stanescu *et al.*, 2015; Tiyaboonchai *et al.*, 2017). This patient has a heterozygous 4 base pair duplication (c.dup606-609) in exon 2 of *GATA6* that leads to a premature STOP codon and a truncated protein (Figure 3-2 A)

(Stanescu *et al.*, 2015). This mutation was previously corrected using CRISPR-CAS9 and used as a genetically matched control (denoted as $iPS^{+/+}$) (Tiyaboonchai *et al.*, 2017). We also generated an identical mutation in an ESC line (denoted as $Mel^{+/mut}$) to compare the effect of this mutation on two different genetic backgrounds (Figure 3-2 A, B and Table 3-1). We used flow cytometry to measure GATA6 protein levels compared to isotype controls and to measure PDX1+ and NKX6.1+ cells at the DE, PFG and PP stages of the differentiation. As expected, we found that all the lines differentiated normally to the PFG stage marked by the percent PDX1+ cells (Figure 3-1 B). By gating on PDX1+ cells at the PP stage, we identified that the mutant cells lines ($iPS^{+/mut}$ and $Mel^{+/mut}$) generated fewer PDX1+/NKX6.1+ PP cells compared to their genetically identical wild type controls ($iPS^{+/+}$ and $Mel^{+/+}$) (Figure 3-1 C, D and Figure 3-2 D). The mutant cells lines also had lower GATA6 protein levels at the PFG and PP stages of the differentiation (Figure 3-1 G).

The defect in generating NKX6.1+ cells at the PP stage could be due to multiple reasons, including increased apoptosis or a switch in the cell fate of the PDX1+ cells caused by *GATA6* haploinsufficiency. It has been previously shown that *GATA6* heterozygous mutants cause no significant apoptosis or proliferation defects at the PP stage (Shi *et al.*, 2017). Interestingly, in mouse models of PA, there is extension of the SOX2+ stomach and CDX2+ intestinal domains into the pancreatic domain (Xuan and Sussel, 2016). To identify any switches in the cell fate at the PP stage caused by the *GATA6* mutation, we measured SOX2+ and CDX2+ cells in the PDX1+ population (Figure 3-2 D and E). We found that the reduced GATA6 levels in the mutant lines at PFG and PP stages leads to a decrease in the NKX6.1+ cells and an increase in SOX2+

cells at the PP stage (Figure 3-1 C, D and E). We detected no SOX2 or CDX2 expression at the PFG stage. We did not detect CDX2 protein during any stage of the differentiation (Figure 3-2 E).

Figure 3-1: *GATA6* haploinsufficiency leads to a pancreas progenitor defect and a switch in cell fate

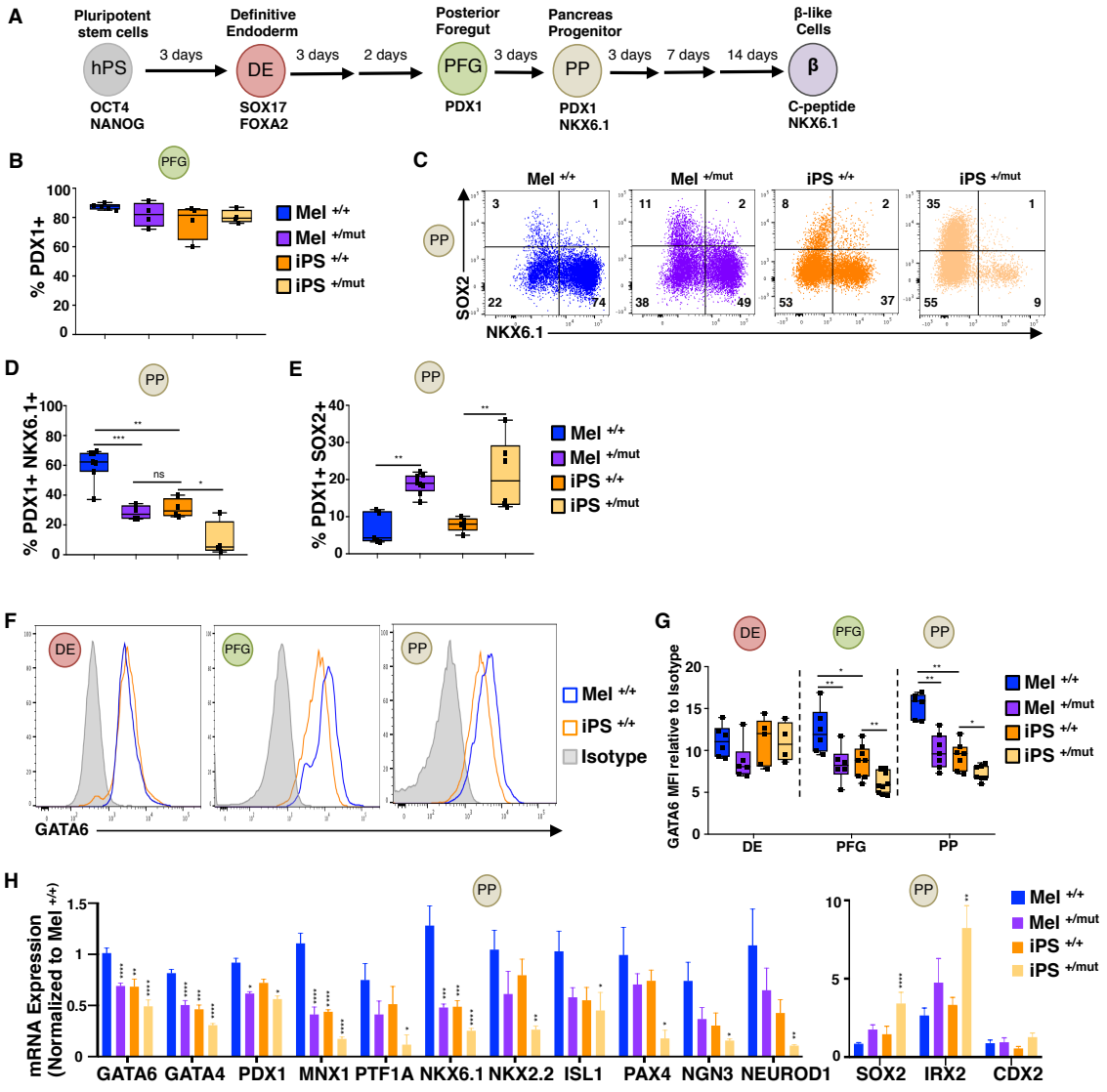


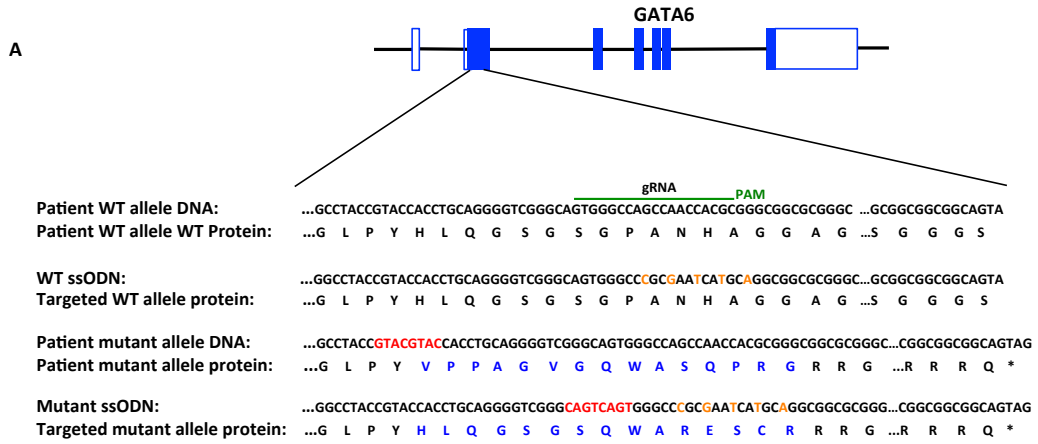
Figure 3-1: *GATA6* haploinsufficiency leads to a pancreas progenitor defect and a switch in cell fate

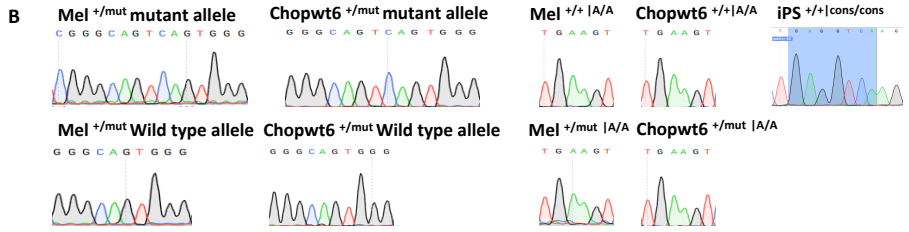
(A) Schematic of pancreas differentiation protocol from hPSCs. The cell types and key markers at each stage are shown. hPS, undifferentiated hPSC stage; DE, definitive endoderm stage; PFG, posterior foregut stage; PP, pancreas progenitor stage; β, β-like cells stage.
(B-H) Data from $Mel^{+/+}$, $Mel^{+/mut}$, $iPS^{+/+}$ and $iPS^{+/mut}$ cell lines.
(B) Flow cytometry quantification of %PDX1+ cells at the PFG stage.
(C) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining gated on PDX1+ cells at the PP stage.
(D-E) Flow cytometry quantification at the PP stage. (D) %PDX1+/NKX6.1+ double positive cells and (E) %PDX1+/SOX2+ double positive cells.
(F) Representative flow cytometry histograms for GATA6 compared to Isotype control at the DE, PFG and PP stages for the $Mel^{+/+}$ and $iPS^{+/+}$ cell lines.
(G) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the DE, PFG and PP stages.
(H) qRT-PCR analysis of key pancreatic and stomach development genes relative to the housekeeping gene TBP and normalized to $Mel^{+/+}$ at the PP stage (n = 4).
All data represented as Mean +/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

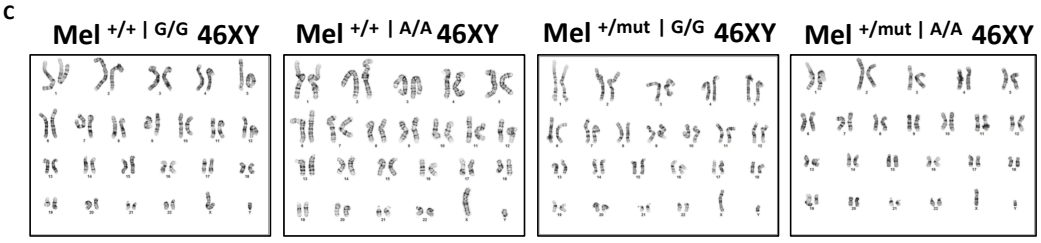
A second interesting observation from these studies was that the patient $iPS^{+/+}$ line generated PDX1+/NKX6.1+ PP cells with far less efficiency compared to the $Mel^{+/+}$ line (31 +/- 3 % vs. 60 +/- 4 %). In fact, the $iPS^{+/+}$ line had a similar efficiency to the $Mel^{+/mut}$ line (28 +/- 2 %) and the $iPS^{+/mut}$ line had much lower efficiency (10 +/- 6 %) (Figure 3-1 C, D and Table 3-2). While it is not surprising that different cell lines have different baseline efficiencies of differentiation, we found that GATA6 protein levels in the $iPS^{+/+}$ and $Mel^{+/+}$ lines were identical at the DE stage. However, at the PFG and PP stages, GATA6 protein levels in the $iPS^{+/+}$ line was statistically lower by ~30% compared to the $Mel^{+/+}$ line (Figure 3-1 F, G and Table 3-3). Similar to the impacts on PP development, GATA6 protein levels were comparable between the $iPS^{+/+}$ and the $Mel^{+/mut}$ lines with the $iPS^{+/mut}$ line being much lower (Figure 3-1 G and Table 3-2). We also measured RNA levels of *GATA4*, a gene that is regulated by GATA6 specifically during human pancreas

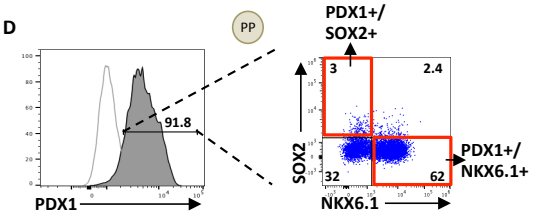
development, and found a reduction in levels similar to *GATA6* (Figure 3-1 H) (Lorberbaum and Sussel, 2017; Tiyaboonchai *et al.*, 2017). RNA expression of *GATA6, NKX6.1 and PDX1* showed the same results as the protein expression of these genes (Figure 3-1 H). Other pancreas specific genes such as *MNX1, PTF1A, NKX2.2, ISL1, NGN3* and *NEUROD1* were also found to be downregulated in a similar manner to GATA6. *SOX2* and *IRX2*, genes involved in stomach development, were found to be upregulated with lower levels of GATA6 (Figure 3-1 H). These data suggest that the patient PA iPS cell line may carry an additional disease modifier that regulates GATA6 expression specifically during pancreas specification.

Figure 3-2: Generating Mel1 ES cell lines









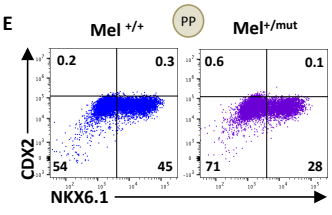


Figure 3-2: Generating Mel1 ES cell lines

(A) CRISPR-CAS9 strategy to generate the GATA6 heterozygous mutation in the Mel ES cell background. The green line marks the guide RNA sequence used. The nucleotides highlighted in orange represent silent mutations incorporated into the ssODN. The nucleotides in red represent the 4 base pair duplication. The amino acids highlighted in blue represent the 15 amino acids that are different between the Mel+/mut and the iPS+/mut cell lines. The * represents the premature STOP codon.
(B) Sequencing reads confirming the introduction of the heterozygous coding region mutations and minor allele variant of rs12953985 in isogenic cell lines.
(C) Karyotypes for the $Mel^{+/+ \mid G/G}$, $Mel^{+/+ \mid A/A}$, $Mel^{+/mut \mid G/G}$, $Mel^{+/mut \mid A/A}$ cell lines.
(D) Flow cytometry gating strategy for generating the NKX6.1+ and SOX2+ dot plots, PDX1+/ NKX6.1+ and the PDX1/ SOX2+ graphs.
(E) Flow cytometry dot plots for NKX6.1 and CDX2 co-staining at the PP stage for the $Mel^{+/+}$ and $Mel^{+/mut}$ cell lines.

3.2.2 The PA patient iPSCs have the minor allele variant of SNP rs12953985

The finding that GATA6 expression was lower in PA patient-derived pancreatic progenitors even when the coding mutation was corrected was particularly interesting because not all patients with *GATA6* heterozygous mutations have PA. All previously described *GATA6* mutations have been identified either by exome sequencing or by amplifying the exons, exon/intron boundaries and small portions of 3' and 5' UTR's of *GATA6* followed by Sanger sequencing. Therefore, we hypothesized that a variant in a non-coding regulatory region of *GATA6* could act in conjunction with a coding mutation resulting in a stronger phenotype. To identify active regulatory regions around *GATA6*, we used a combination of three strategies. First, we identified regions enriched for H3K4Me1 at the PP stage in a 200kb window surrounding *GATA6* using previously published data (Figure 3-3 A) (Weedon, Ines Cebola, *et al.*, 2014). Second, using previously published datasets we identified transcription factor binding regions for PDX1,

HNF6 and FOXA2 that are known regulators of pancreas development at the PP stage (Figure 3-3 A) (Weedon, Ines Cebola, *et al.*, 2014; Teo *et al.*, 2015). Third, we analyzed H3K4Me1, H3K27Ac and DNAse hypersensitivity sites from the Roadmap Epigenetics Consortium (REMC) in fetal human tissues that arise from the gut tube, proximal to the pancreas, such as the small and large intestine, stomach and adult liver (Figure 3-3 B) (Consortium *et al.*, 2015). These strategies narrowed our search to regions enriched for H3K4Me1 and that bind at least two of the three transcription factors.

Upon sequencing these regions, the only difference we discovered between the patient iPSC line and the Mel1 line was within a 3' region ~8kb downstream of *GATA6* (designated R5) (Figure 3-3 C). The patient iPS cell line was homozygous for the minor allele A of the SNP rs12953985 and the Mel1 line was homozygous for the major allele G variant (Figure 3-3 C). We cloned the entire R5 region into the pGL4.23 vector and performed enhancer luciferase assays at the DE and PFG stages of differentiation. This region showed luciferase activity over the empty pGL4.23 vector control at the PFG stage, but not the DE stage (Figure 3-3 D). Recently published ATAC-seq data also show that the chromatin in region R5 is accessible only during the PFG stage but not during earlier stages of the differentiation (Figure 3-3 E) (Lee *et al.*, 2019). Furthermore, the homozygous minor allele variant of rs12953985 correlated with lower *GATA6* RNA levels in the pituitary and testis when analyzed on GtexEQTL (Figure 3-3 F). Interestingly, this region is conserved in primates but not in rodents, which is consistent with the differences in regulation of *GATA6* during human pancreas development compared to mice (Figure 3-3 H).

Figure 3-3: Identifying regulatory regions of GATA6

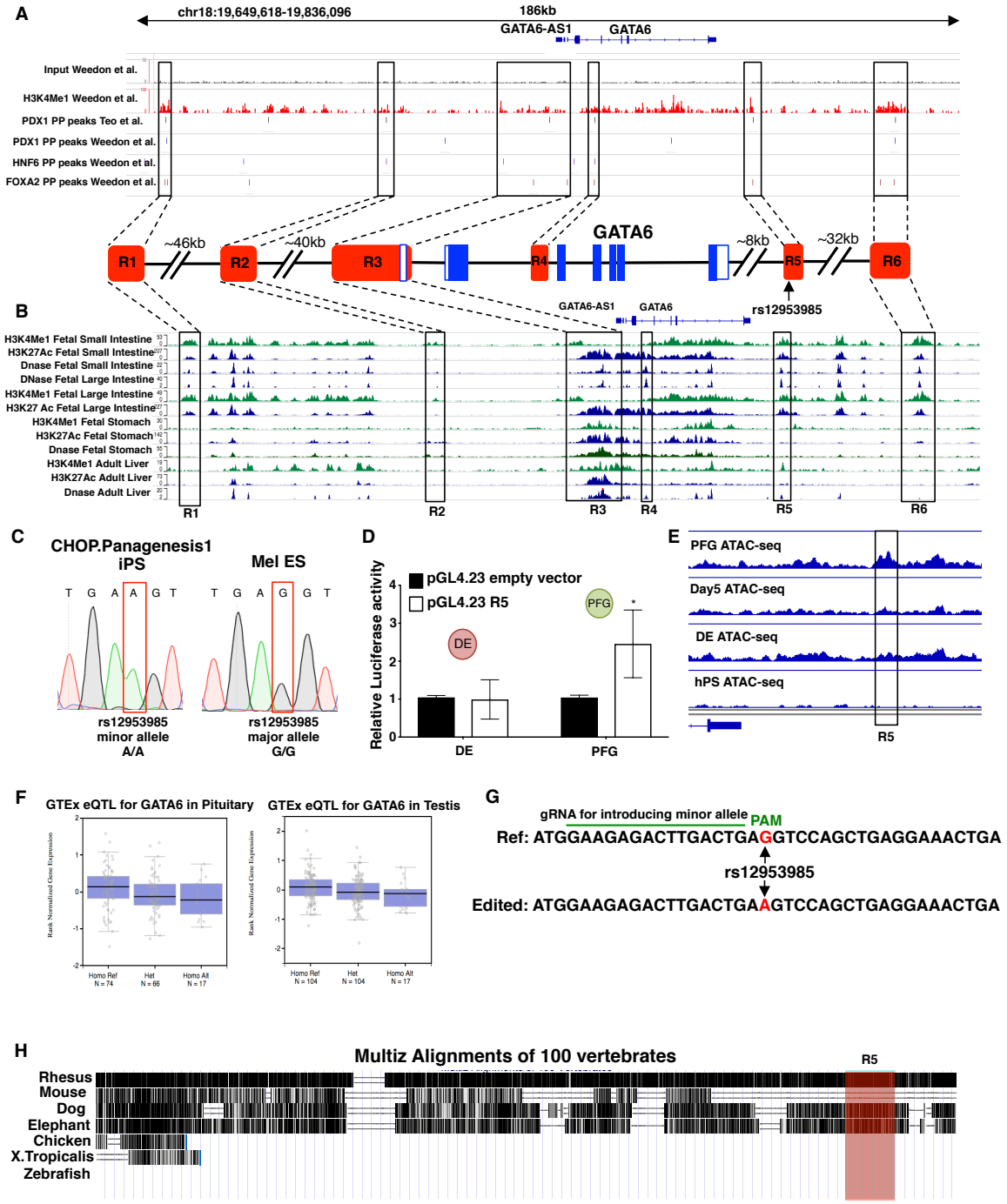


Figure 3-3: Identifying regulatory regions of GATA6

(A) Integrative Genomics Viewer (IGV) snapshot of chr18: 19649618-19836096 on hg19. Bedgraph tracks for Input, H3K4Me1 and MACS peaks for PDX1, HNF6 and FOXA2 are plotted. Below: Schematic of the GATA6 locus including exons (solid blue boxes), Untranslated regions (hollow blue boxes), regions with H3K4Me1 enrichment with at least 2 transcription factors bound (red boxes labeled R1 to R6).
(B) WashU epigenome browser snapshot of chr18: 19649618-19836096 on hg19. Bigwig tracks for H3K4Me1, H3K27Ac and DNase hypersensitivity sites from fetal small intestine, fetal large intestine, fetal stomach and adult liver are shown. The black boxes indicate regions R1-R6.
(C) Sequencing data from the CHOP.Panagenesis1 patient iPS line and Mel ES cell line for rs12953985.
(D) Enhancer luciferase assays at the DE and PFG stages for R5. All values are nomalized to renilla luciferase to control for transfection efficiency and then to the pGL4.23 empty vector transfection. Unpaired t-test with two-tailed distribution was used to determine the significance.
(E) IGV snapshot of R5 showing ATAC-seq tracks from Lee et al. at hPS, DE, Day 5 and PFG stages of the differentiation.
(F) GTEx eQTL plots for rs12953985 variant effect on GATA6 expression in pituitary and testis.
(G) CRISPR-CAS9 strategy to generate the minor allele variant of rs12953985 in the $Mel^{+/+ \mid G/G}$ and $Mel^{+/mut \mid G/G}$ background. The green line marks the guide RNA sequence used.
(H) Multiz Alignments from UCSC genome browser for human, rhesus, mouse, dog, elephant, chicken, x.tropicalis and zebrafish. The red highlight represents region R5.

3.2.3 The minor allele frequency of the SNP rs12953985 is enriched in PA patients

To confirm this finding in a larger PA patient cohort, we first tested the SNP in 4 patients with PA caused by *GATA6* mutations referred to the University of Chicago monogenic diabetes registry for genetic testing (Figure 3-3 A and Table 3-4). We sequenced the DNA of the 4 PA patients for rs12953985 and found that all carried the minor allele; 2 were heterozygous A/G and 2 were homozygous A/A for the minor allele variant. The overall allele frequency in this group was 75% [95%CI=24.5%] (Figure 3-4 A and Table 3-4). We also sequenced eight wild type hPSC lines for rs12953985 and six of them were homozygous G/G for the major allele while only two were heterozygous A/G (Figure 3-5 B). We found no homozygous A/A minor allele variants. These findings

suggest that rs12953985 may be a disease modifying SNP that can regulate GATA6 expression.

Figure 3-4: The minor allele frequency of SNP rs12953985 is higher in PA patients

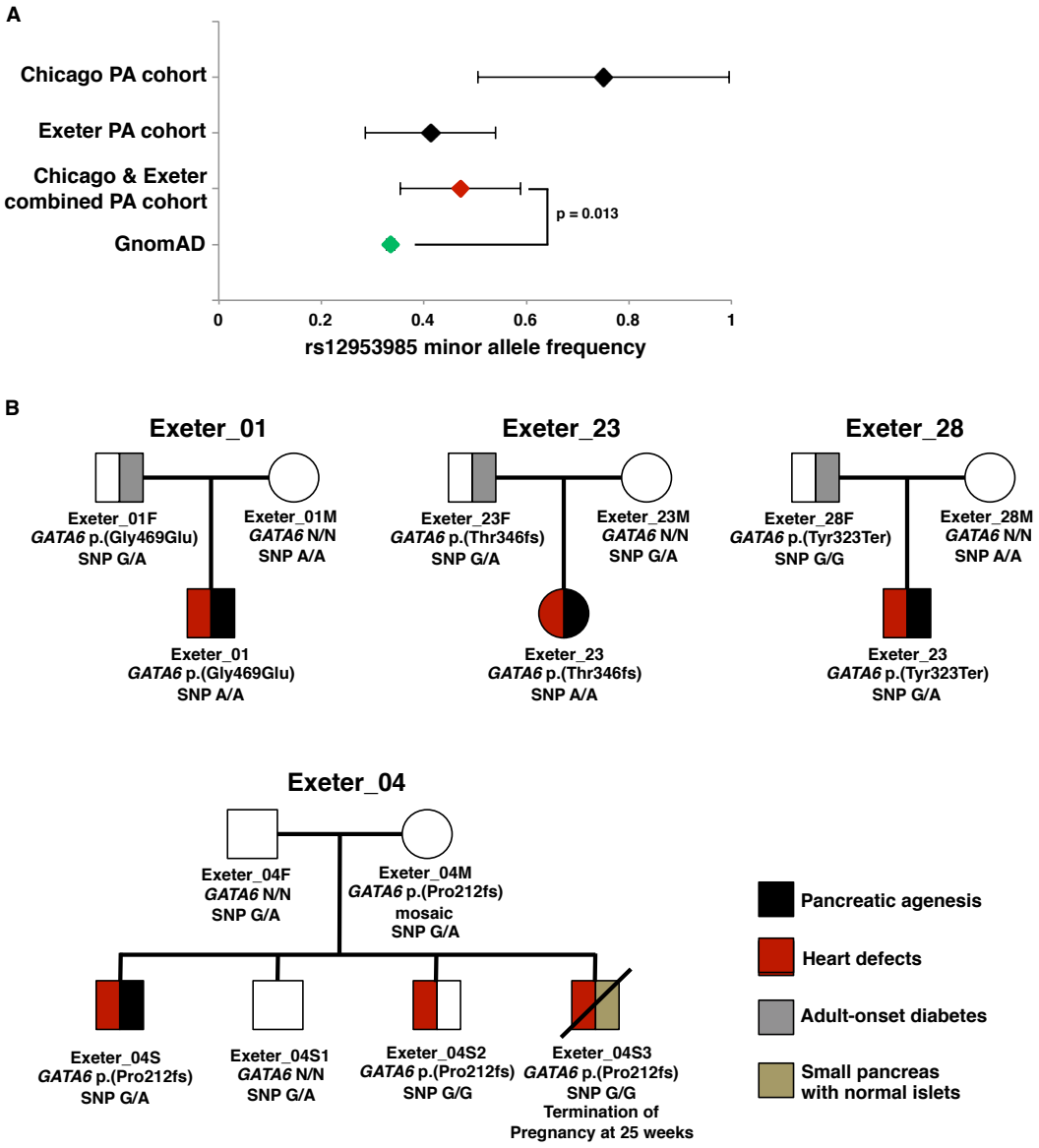


Figure 3-4: **The minor allele frequency of SNP rs12953985 is higher in PA patients**

(A) rs12953985 minor allele frequency in individuals with PA from the Chicago, Exeter and combined cohorts compared to GnomAD European database. Plots show the frequency of the minor allele A and the 95% confidence intervals.
(B) 4 pedigrees from the Exeter cohort, where the GATA6 mutation was inherited from a parent with a mild or no pancreatic phenotype and the minor allele variant of the SNP was inherited in trans from the other parent. Colors indicate phenotype of the patients as follows: Black, PA; Red, Heart Defects; Grey, Adult-onset diabetes; Olive, Small pancreas with normal islets.

Figure 3-5: Sequencing patient and wild type lines for rs12953985 variant

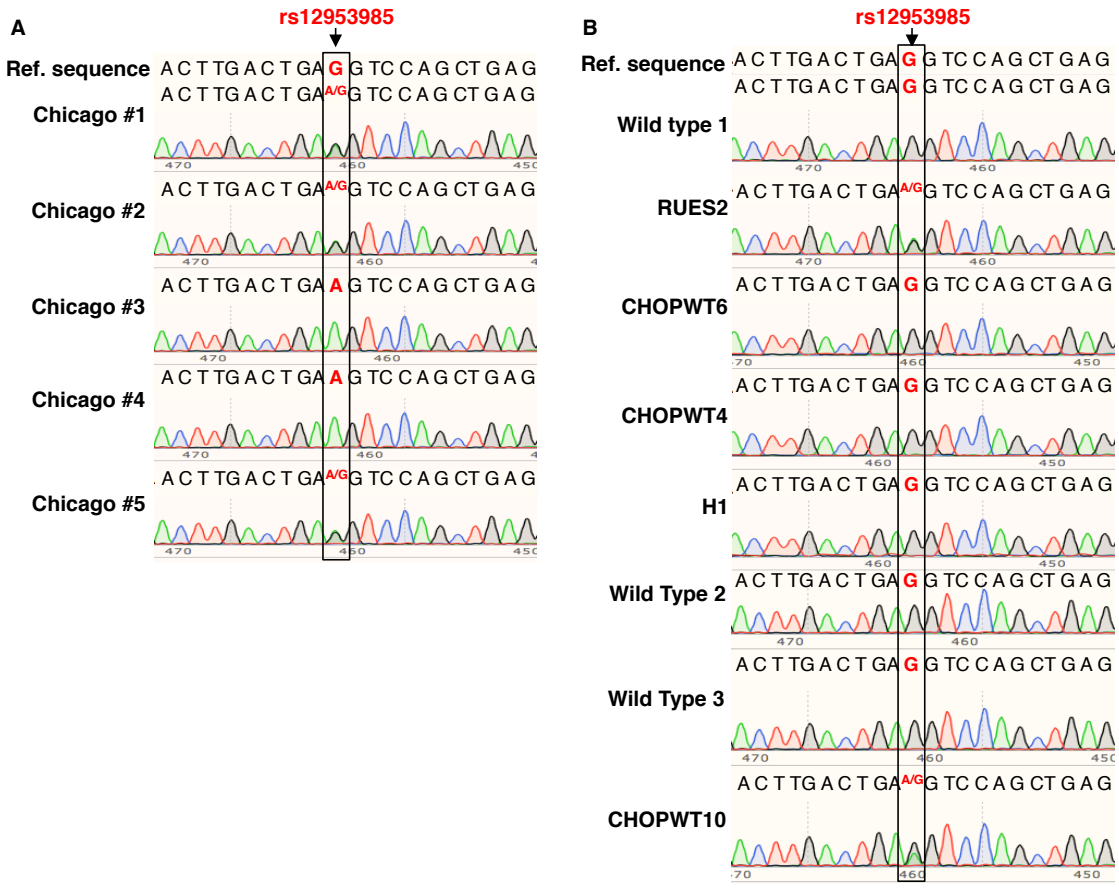


Figure 3-5: Sequencing patient and wild type lines for rs12953985 variant

(A) Reference sequence and sequencing results for 5 PA patient samples from the Chicago cohort described in Table S5.
(B) Reference sequence and sequencing results for 8 wild type hPSC lines.

We sought to replicate these findings in a cohort of 36 patients of European ancestry (28 with PA, 2 with transient neonatal diabetes, 4 with diabetes diagnosed in adulthood, and 2 with congenital heart defects but no diabetes) heterozygous for *GATA6* mutations referred to the Exeter Molecular Genetics laboratory (Table 3-4). The minor allele, A, was again found to be more frequent among patients with PA compared to the non-PA group (allele frequency 42.3% [95%CI=13%] vs 22.2% [95%CI=19.2%]) (Figure 3-4 A and Table 3-4). The A allele was present in 25/33 individuals with *GATA6* PA tested in the two cohorts for a cumulative allele frequency of 48.5% [95%CI=11.9%] (Figure 3-4 A and Table 3-4). This is higher than the allele frequency among patients with the *GATA6* mutation without PA (27.8% [95%CI=20.7%]) and the frequency among Europeans in the GnomAD database (33.5% [95%CI=0.75%], ChiSquare p= 0.013) (Figure 3-4 A).

Phasing of the *GATA6* mutation and the A allele of the SNP in 10 families (8 homozygotes AA and two heterozygotes GA with an inherited *GATA6* mutation) showed that the A variant was on the opposite allele of the mutation in 7 cases with PA and only one case without PA (diagnosed with diabetes at 12 years) (Table 3-4). Interestingly, in four pedigrees with multiple individuals harboring GATA6 mutations, the A allele was *in trans* with the coding mutation in all the probands with PA while 4 family members without PA either carried the A allele *in cis* with the coding mutation or were homozygous GG (Figure 3-4 B). The most recent pregnancy for Family Exeter_04 carried a pathogenic *GATA6* mutation and was homozygous GG for the SNP. Severe congenital heart defects were detected prenatally leading to a termination of pregnancy. At post-mortem the fetus was found to have a small pancreas with normal islet architecture (Yau *et al.*, 2017). The severity of the pancreatic phenotype in this individual

is therefore impossible to assess. The Odds Ratio for carrying the A allele *in trans* with the mutation suggested that this allele strongly increases the risk of pancreatic agenesis [OR=7.78, 95%CI=0.8-76.1], however replication in larger PA and non-PA cohorts is needed in order to confirm this result. These results suggest the SNP A allele is associated with PA and that it is usually found *in trans* with the coding mutation. These data are consistent with the hypothesis that the SNP A allele when present on the only functional coding allele of GATA6 will lower expression below a threshold required for proper pancreas development.

3.2.4 The minor allele variant of rs12953985 lowers GATA6 expression during pancreas specification

In order to add functional evidence that this SNP influences GATA6 expression and pancreas development, we introduced the minor allele variant of the SNP, rs12953985, into the $Mel^{+/+}$ and $Mel^{+/mut}$ cell lines using CRISPR-CAS9 technology (Figure 3-3 G). We differentiated the four genetically matched cell lines, $Mel^{+/+ \mid G/G}$, $Mel^{+/mut \mid G/G}$, $Mel^{+/+ \mid A/A}$ and $Mel^{+/mut \mid A/A}$ (Table 3-1), to measure GATA6 protein levels at the DE, PFG and PP stages and to measure the efficiency of PDX1+/NKX6.1+ cell generation at the PP stage. The $Mel^{+/+ \mid G/G}$ had the highest GATA6 expression at the PFG and PP stages (Figures 3-6 A, B and Table 3-3).

Figure 3-6: The minor allele variant of rs12953985 affects pancreas development

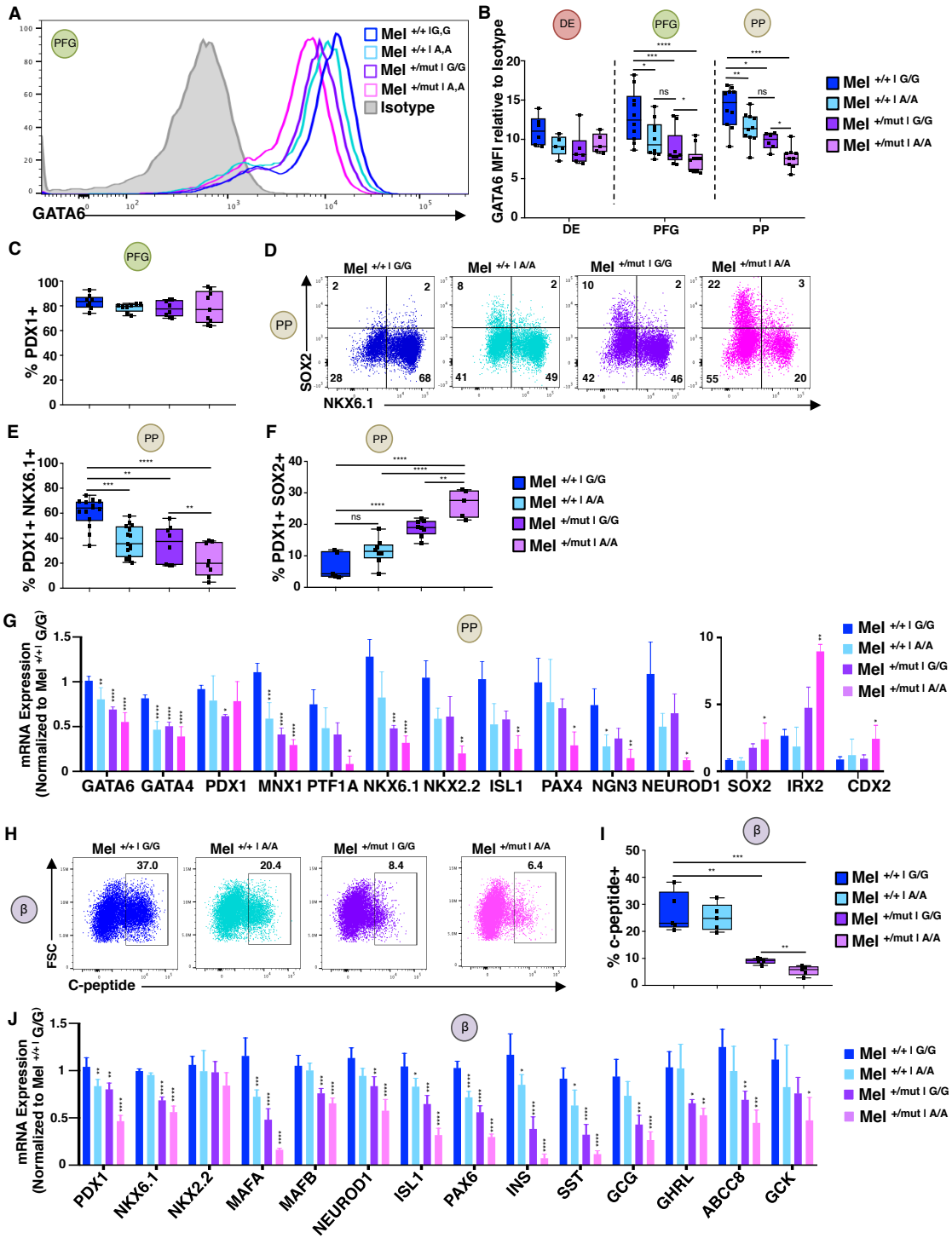


Figure 3-6: The minor allele variant of rs12953985 affects pancreas development

(A-J) Data from Mel$^{+/+ \mid G/G}$, Mel$^{+/+ \mid A/A}$, Mel$^{+/mut \mid G/G}$, Mel$^{+/mut \mid A/A}$ cell lines.
(A) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG stage.
(B) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the DE, PFG and PP stages.
(C) Flow cytometry quantification of %PDX1+ cells at the PFG stage.
(D) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining at the PP stage.
(E-F) Flow cytometry quantification at the PP stage. (E) %PDX1+/NKX6.1+ double positive cells and (F) %PDX1+/SOX2+ double positive cells.
(G) qRT-PCR analysis of key pancreatic and stomach development genes relative to the housekeeping gene TBP and normalized to Mel$^{+/+ \mid G/G}$ at the PP stage.
(H) Representative flow cytometry dot plot for C-peptide and forward scatter (FSC) at the β stage.
(I) Quantification of %C-peptide+ cells at the β-like stage. Unpaired two-tailed t-tests were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001.
(J) qRT-PCR analysis of key pancreatic β cell and islet signature genes relative to the housekeeping gene TBP and normalized to Mel$^{+/+ \mid G/G}$ at the β-like stage.
All data represented as Mean +/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

Both, the Mel$^{+/mut \mid G/G}$ and Mel$^{+/+ \mid A/A}$ lines, showed ~25% reduction in GATA6 expression while the Mel$^{+/mut \mid A/A}$ had the lowest GATA6 expression, ~50% reduction, compared to the Mel$^{+/+ \mid G/G}$ line at the PFG and PP stages (Figures 3-6 A, B and Table 3-3). This reduction in GATA6 protein expression was also verified by western blot analyses at the PFG and PP stages (Figure 3-7 A). The consequence of this reduction in GATA6 expression was seen at the PP stage in the efficiency of generating NKX6.1+ cells or SOX2+ cells (Figures 3-6 C, D, E, F and Table 3-4). The Mel$^{+/+ \mid G/G}$ line had the highest efficiency of generating NKX6.1+ cells and the lowest efficiency of generating SOX2+ cells at the PP stage (60 +/- 3 % and 6 +/- 2 % respectively). The Mel$^{+/+ \mid A/A}$ and the Mel$^{+/mut \mid G/G}$ lines had a lower efficiency of generating NKX6.1+ cells and concurrently an increase in SOX2+ cells. The Mel$^{+/mut \mid A/A}$ line generated the fewest NKX6.1+ cells and the most SOX2+ cells at the PP stage (22 +/- 4 % and 26 +/- 2 % respectively). These

results were also verified using immunofluorescence (Figure 3-7 B). RNA expression of key pancreatic genes *GATA4, SOX2, NKX6.1, PDX1*, *MNX1, PTF1A, NKX2.2, ISL1, NGN3* and *NEUROD1,* were lower in the lines with lower *GATA6* expression (Figure 3-6 G). Conversely, *SOX2* and *IRX2* RNA expression were higher in these lines suggesting a switch in the cell fate of PDX1+ cells at the PP stage (Figure 3-6 G). Previous studies have shown that *GATA6* heterozygous mutations lead to defects in generating C-peptide positive cells at the β-like cell stage (Shi *et al.*, 2017; Tiyaboonchai *et al.*, 2017). To confirm this, all four lines were differentiated to the β-like cell stage and CPEP+ cells were measured. As expected, CPEP+ cells from the $Mel^{+/mut\ |\ G/G}$ line were reduced compared to the CPEP+ cells from the $Mel^{+/+\ |\ G/G}$ line (9 +/- 1% vs. 33 +/- 11%)(Figures 3-6 H and I). The minor allele variant line harboring the coding mutation ($Mel^{+/mut\ |\ A/A}$) demonstrated a statistically significant decrease in CPEP+ cells compared to its major allele variant counterpart ($Mel^{+/mut\ |\ G/G}$)(Figures 3-6 H and I). This decrease in CPEP+ cells was more pronounced in the CHOPWT6 genetic background (Figures 3-8 I and J). We also detected a decrease in the percentage of SST+ and GCG+ cells at this stage (Figures 3-7 C and D). The expression of key β cell genes, islet hormone genes and genes related to β cell functionality were also lower in the lines with lower GATA6 levels (Figure 3-6 J). These experiments were also replicated using cells of a different genetic background (CHOPWT6 iPS) with similar results (Figure 3-8).

Figure 3-7: Testing the effect of rs12953985 on isogenic Mel ES cell lines

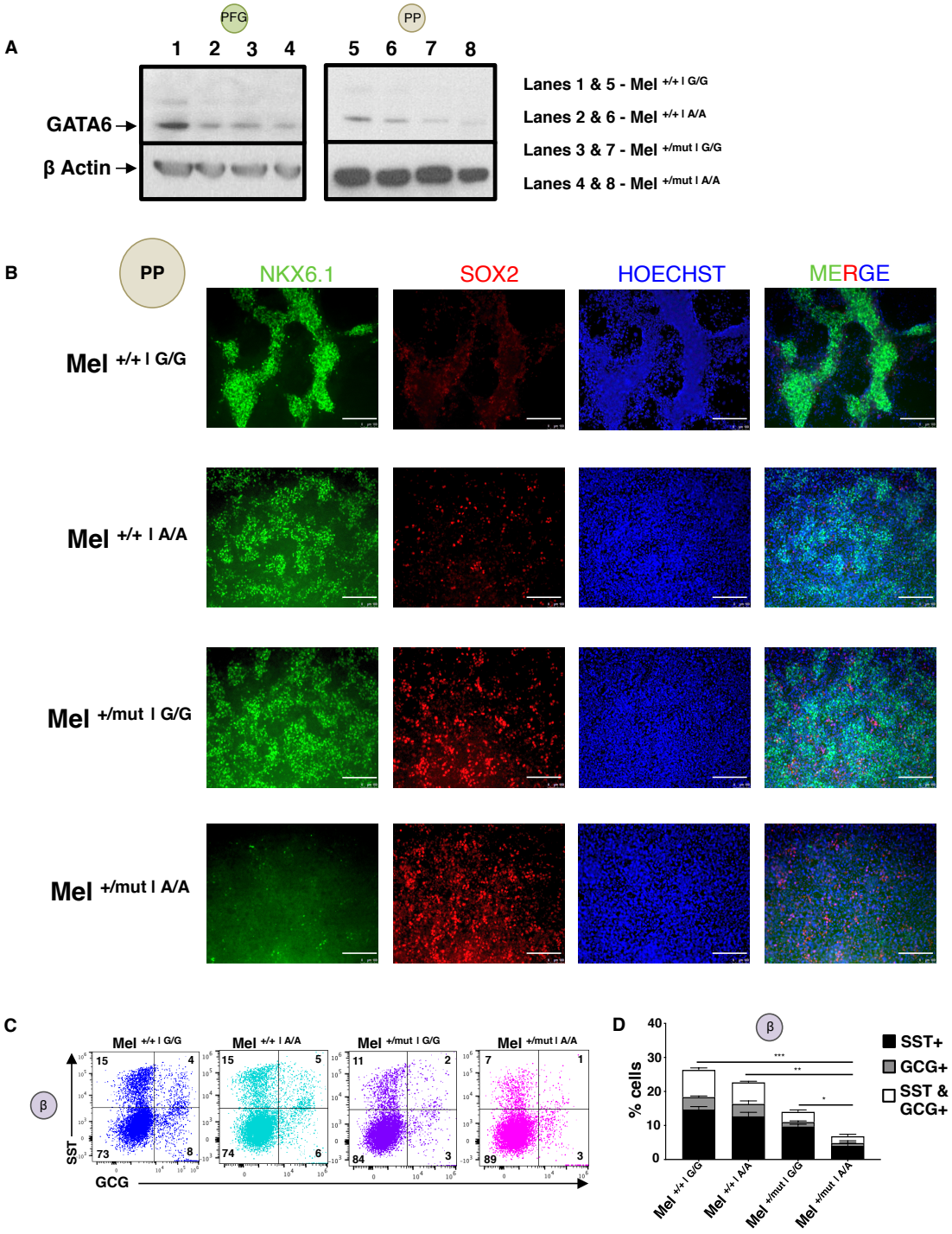


Figure 3-7: Testing the effect of rs12953985 on isogenic Mel ES cell lines

(A) Western blot for GATA6 and loading control β Actin at the PFG and PP stages in the $Mel^{+/+ | G/G}$, $Mel^{+/+ | A/A}$, $Mel^{+/mut | G/G}$, $Mel^{+/mut | A/A}$ cell lines. Lanes 1 and 5 represent the $Mel^{+/+ | G/G}$ at the PFG and PP stages respectively; Lanes 2 and 6 represent the $Mel^{+/+ | A/A}$ at the PFG and PP stages respectively; Lanes 3 and 7 represent the $Mel^{+/mut | G/G}$ at the PFG and PP stages respectively; Lanes 4 and 8 represent the $Mel^{+/mut | A/A}$ at the PFG and PP stages respectively.
(B) Immunofluorescence images at the PP stage for NKX6.1 in green, SOX2 in red and HOECHST DNA stain in blue. Scale bar 200um.
(C) Representative flow cytometry dot plot for SST and GCG co-staining at the β-like stage.
(D) Flow cytometry quantification of SST+, GCG+ and SST+/ GCG+ double positive cells at β-like stage.
All data is represented as Mean+/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

3.2.5 The minor allele variant of rs12953985 disrupts RORα binding

To determine if the minor allele variant of rs12953985 disrupted any transcription factor binding sites, we used CIS-BP (Weirauch *et al.*, 2014). We identified a RORα binding domain that could potentially be disrupted by the G>A variant (Figure 3-9 A). Recent bioinformatics data have suggested a possible role for RORα in human pancreas development (Jennings *et al.*, 2017). To detect RORα expression during the *in vitro* differentiation protocol, qRT-PCR and immunofluorescence were performed at the DE, PFG and PP stages (Figure 3-9 B and C). Robust RNA expression was detected at the DE stage with slightly lower expression at the PFG and PP stages (Figure 3-9 B).

Figure 3-8: Testing the effect of rs12953985 on isogenic CHOPWT6 iPS cell lines

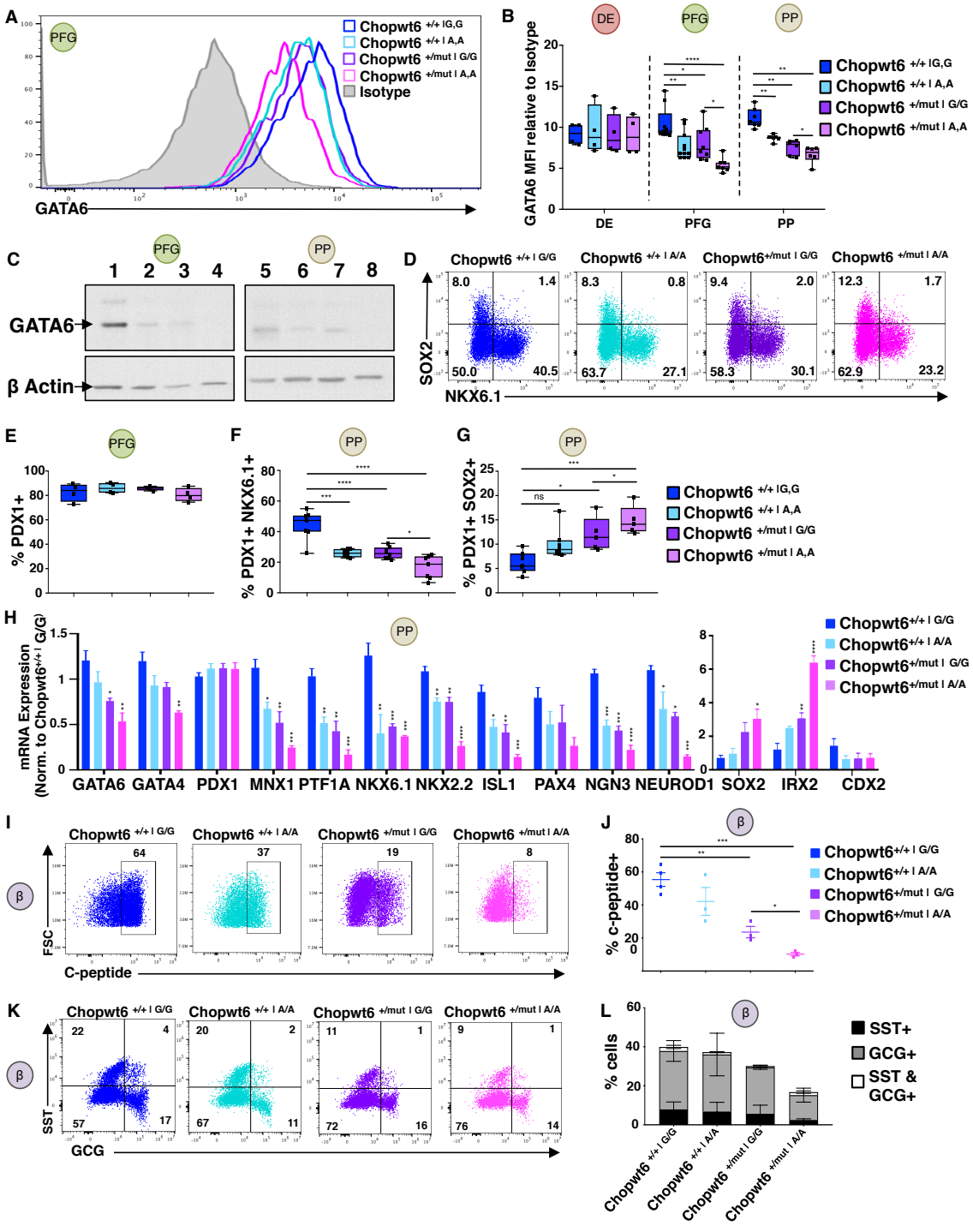


Figure 3-8: Testing the effect of rs12953985 on isogenic CHOPWT6 iPS cell lines

(A-L) Data from Chopwt6$^{+/+ \mid G/G}$, Chopwt6$^{+/+ \mid A/A}$, Chopwt6$^{+/mut \mid G/G}$, Chopwt6$^{+/mut \mid A/A}$ cell lines.
(A) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG stage.
(B) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the DE, PFG and PP stages.
(C) Western blot for GATA6 and loading control β Actin at the PFG and PP stages. Lanes 1 and 5 represent the Chopwt6$^{+/+ \mid G/G}$ at the PFG and PP stages respectively; Lanes 2 and 6 represent the Chopwt6$^{+/+ \mid A/A}$ at the PFG and PP stages respectively; Lanes 3 and 7 represent the Chopwt6$^{+/mut \mid G/G}$ at the PFG and PP stages respectively; Lanes 4 and 8 represent the Chopwt6$^{+/mut \mid A/A}$ at the PFG and PP stages respectively.
(D) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining at the PP stage.
(E-G) Flow cytometry quantification of (E) %PDX1+ cells at the PFG stage (F) %PDX1+/NKX6.1+ double positive cells at the PP stage and (G) %PDX1+/SOX2+ double positive cells at the PP stage.
(H) qRT-PCR analysis of key pancreatic and stomach development genes relative to the housekeeping gene TBP and normalized to Chopwt6$^{+/+ \mid G/G}$ at the PP stage.
(I) Representative flow cytometry dot plot for C-peptide and forward scatter (FSC) at the β stage.
(J) Quantification of %C-peptide+ cells at the β-like stage. Unpaired two-tailed t-tests were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001.
(K) Representative flow cytometry dot plot for SST and GCG co-staining at the β-like stage.
(L) Flow cytometry quantification of SST+, GCG+ and SST+/ GCG+ double positive cells at β-like stage.
All data is represented as Mean+/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

We also did not detect expression of other ROR family members, *RORB* and *RORC* (Figure 3-9 B and data not shown). Immunofluorescence showed RORα protein expression at both DE and PFG stages (Figure 3-9 C). To determine RORα binding efficiency in the presence of the G or A variant, ChIP-qPCR was performed on all four lines at the PFG stage (Figure 3-9 D). The lines with the minor allele variant bound RORα with lower efficiency compared to the major allele variant controls. At rs12953985, we detected RORα binding in both Mel$^{+/+ \mid G/G}$ and the Mel$^{+/mut \mid G/G}$ lines over IgG control (2.9 +/- 0.2 and 2.8 +/- 0.6 respectively). However, in the Mel$^{+/+ \mid A/A}$ and the Mel$^{+/mut \mid A/A}$ lines, RORα binding was significantly reduced (1.7 +/- 0.2 and 1.4 +/- 0.4 respectively). RORα binding was unaffected at positive and negative control regions in all four lines

regardless of the GATA6 mutation or rs12953985 variant (Figure 3-9 D). These findings suggest that RORα binds at the 3' regulatory region, R5, specifically in the presence of the major allele G variant of rs12953985.

Figure 3-9: RORα regulates GATA6 during pancreas development

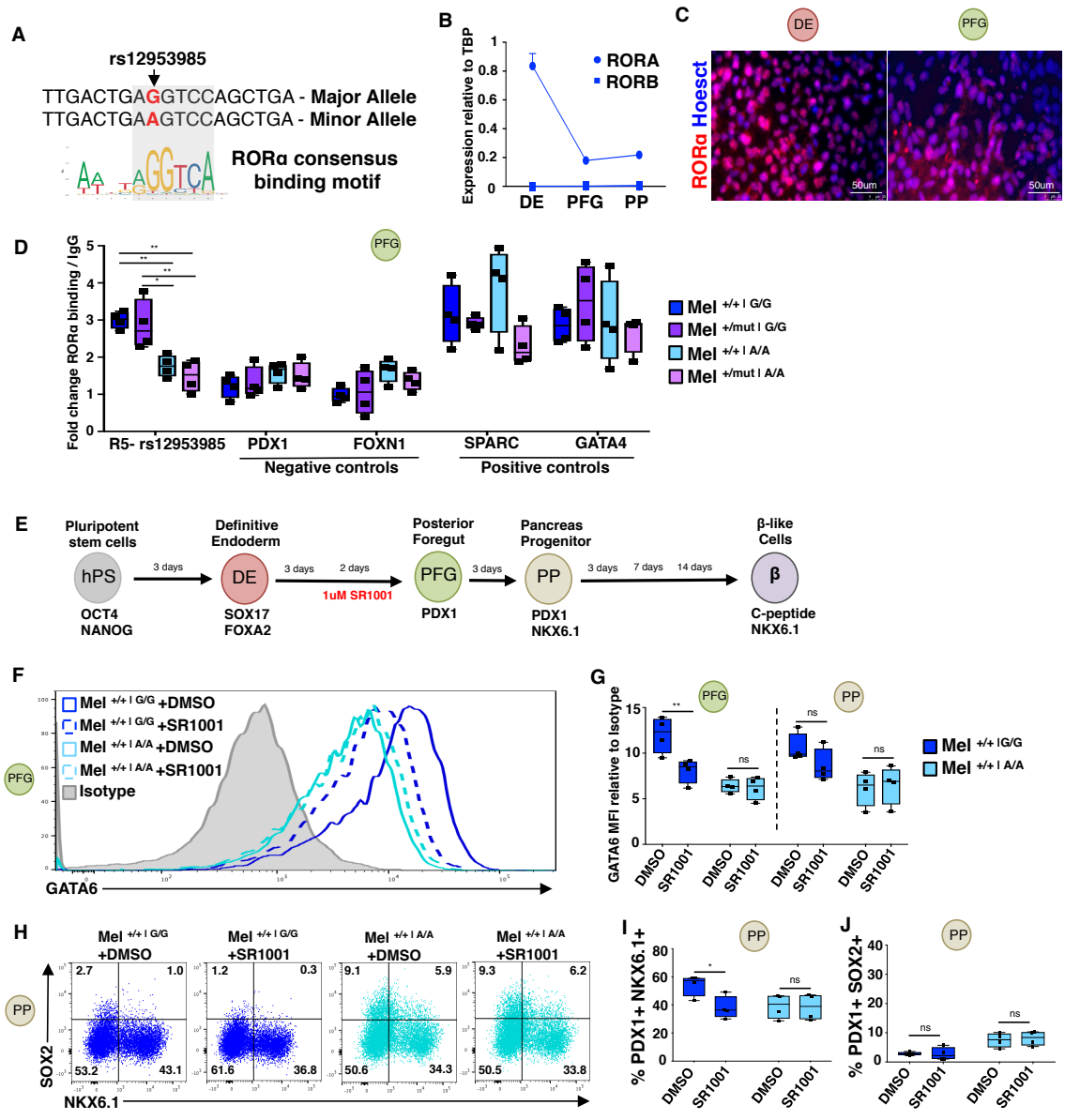


Figure 3-9: RORα regulates GATA6 during pancreas development

(A) Sequence of R5 around rs12953985 (nucleotide in red) and the consensus RORα binding motif from JASPAR.
(B) qRT-PCR analysis of *RORA* and *RORB* relative to the housekeeping gene TBP at the DE, PFG and PP stages in the $Mel^{+/+ | G/G}$ cell line.
(C) Immunofluorescence images for cells stained with RORα (red) and Hoechst (blue) at the DE and PFG stages in the $Mel^{+/+ | G/G}$ cell line.
(D) ChIP-qRTPCR for RORα normalized to IgG control at R5 around rs12953985 and regulatory regions of PDX1, FOXN1, SPARC and GATA4 at the PFG stage for the $Mel^{+/+ | G/G}$, $Mel^{+/+ | A/A}$, $Mel^{+/mut | G/G}$, $Mel^{+/mut | A/A}$ cell lines.
(E) Schematic of pancreas differentiation protocol from hPSCs. 1uM SR1001 was added for 2 days before the PFG stage.
(F-J) Data from $Mel^{+/+ | G/G}$ and $Mel^{+/+ | A/A}$ cell lines treated with DMSO or SR1001.
(F) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG stage.
(G) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the PFG and PP stages.
(H) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining at the PP stage.
(I) Quantification of %PDX1+/NKX6.1+ double positive cells at the PP stage.
(J) Quantification of %PDX1+/SOX2+ double positive cells at the PP stage.
All data represented as Mean +/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

3.2.6 Inhibiting RORα during pancreas specification disrupts pancreas differentiation

Next, we used a known inverse agonist of RORα, SR1001, during PFG specification (Figure 3-9 E). SR1001 binds to the ligand binding domain of RORα and specifically represses activity of genes that RORα activates in the absence of any ligand (Solt and Burris, 2012). We hypothesized that SR1001 should selectively reduce GATA6 expression only in the presence of the major allele G variant of rs12953985. Since the minor allele A variant disrupts RORα binding, SR1001 should have no further effect on reducing GATA6 expression. We added SR1001 for 48 hours prior to the PFG stage and measured GATA6 levels at the PFG and PP stages. We found a significant reduction in GATA6 protein expression at the PFG stage in the line with the G allele while the line

with the A allele was unaffected (Figures 3-9 F, G and Table 3-2). The consequence of this reduction in GATA6 levels at the PFG stage were seen in the efficiency of generating PDX1+/NKX6.1+ cells at the PP stage. The SR1001 treated cultures led to a decrease in NKX6.1+ cells compared to dimethyl sulfoxide (DMSO) controls specifically in the Mel$^{+/+\ |\ G/G}$ line (35 +/- 10 % and 53 +/- 10 % respectively) (Figures 3-9 H, I and Table 3-3). These experiments were replicated with similar results in the CHOPWT6 background (Figure 3-10).

Figure 3-10: RORα regulates GATA6 in isogenic CHOPWT6 iPS cell lines

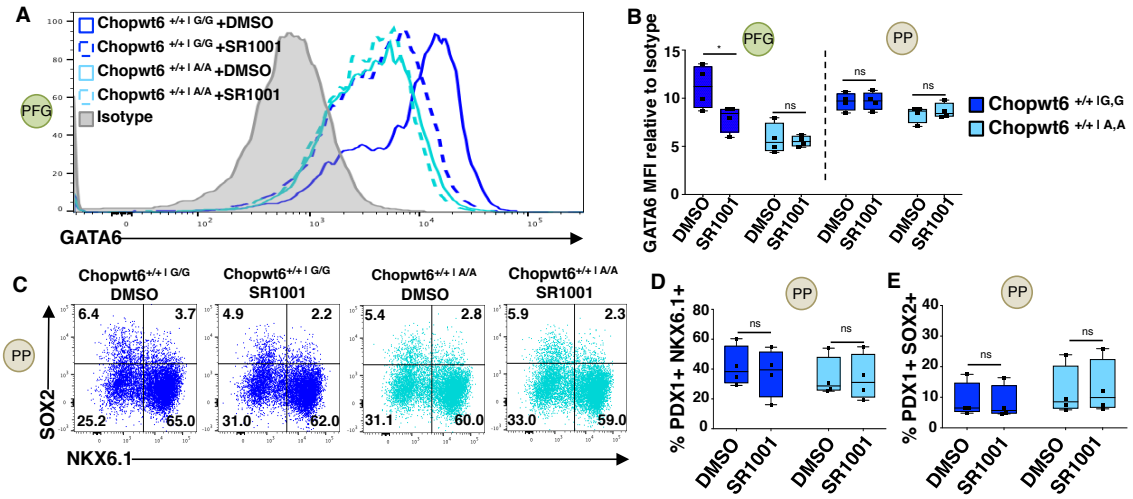


Figure 3-10: RORα regulates GATA6 in isogenic CHOPWT6 iPS cell lines

(A-E) Data from Chopwt6$^{+/+\ |\ G/G}$ and Chopwt6$^{+/+\ |\ A/A}$ cell lines treated with DMSO or SR1001.
(A) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG stage.
(B) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the PFG and PP stages.
(C) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining at the PP stage.
(D) Quantification of %PDX1+/NKX6.1+ double positive cells at the PP stage.
(E) Quantification of %PDX1+/SOX2+ double positive cells at the PP stage.
All data is represented as Mean+/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

We also used siRNA to knockdown RORα specifically during pancreas specification. Again, we hypothesized that the RORα siRNA should selectively reduce GATA6 expression only in the presence of the major allele G variant of rs12953985. RORα or scrambled siRNAs were transfected 48 hours prior to the PFG stage and cells were harvested at the PFG stage (Figure 3-11 A). We measured RORα protein expression using immunofluorescence and RNA expression using qRT-PCR after transfection with RORα or scrambled siRNAs (Figures 3-11 B and C). We found a significant reduction in GATA6 protein expression at the PFG stage in the line with the G alleles while the line with the A alleles were unaffected (Figures 3-11 D and E). Together, these data suggest that modulation of RORα regulates pancreas development via GATA6.

3.2.7 Modifying the minor allele variant to a consensus RORα binding site rescues GATA6 expression and pancreas differentiation

To confirm the effect of RORα on GATA6 expression, a RORα binding site was introduced into the coding mutation corrected PA iPSC line expressing the minor allele variant using CRISPR-CAS9 genome editing. Initial attempts to generate a single A to G base pair change were unsuccessful due to indel formation (data not shown). To bypass this technical hurdle, a consensus RORα motif was used in which two base changes to the original sequence were introduced (Figure 3-12 A) and allowed the generation of modified clones without indels. This line is designated as $iPS^{+/+\ |\ cons/cons}$ (Table 3-1). The $iPS^{+/+\ |\ A/A}$ and the $iPS^{+/+\ |\ cons/cons}$ lines were differentiated and GATA6 levels were measured at the PFG and PP stages.

Figure 3-11: RORα siRNA disrupts GATA6 expression

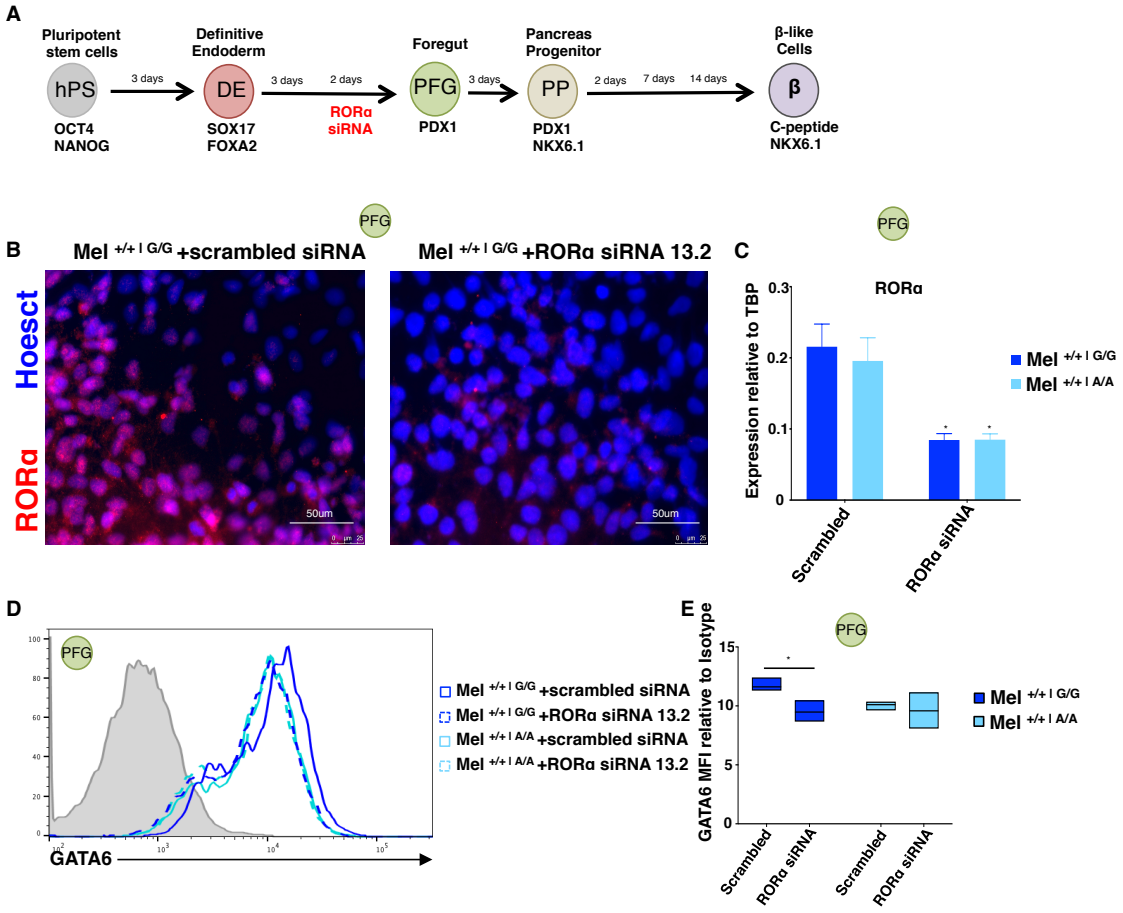


Figure 3-11: RORα siRNA disrupts GATA6 expression during pancreas development

(A) Schematic of pancreas differentiation protocol from hPSCs. 10nM of RORα siRNA or scrambled siRNA was added 2 days prior to the PFG stage.
(B) Immunofluorescence images for cells stained with RORα (red) and Hoechst (blue) at the PFG stages in the Mel$^{+/+\,|\,G/G}$ cell line with RORα siRNA and scrambled siRNA.
(C) qRT-PCR analysis of RORα at the PFG stage in Mel$^{+/+\,|\,G/G}$ and Mel$^{+/+\,|\,A/A}$ cell lines treated with RORα siRNA or scrambled siRNA.
(D) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG stage for the Mel$^{+/+\,|\,G/G}$ and Mel$^{+/+\,|\,A/A}$ cell lines treated with RORα siRNA or scrambled siRNA.
(E) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the PFG and PP stages for the Mel$^{+/+\,|\,G/G}$ and Mel$^{+/+\,|\,A/A}$ cell lines treated with RORα siRNA or scrambled siRNA.
All data is represented as Mean+/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

The introduced RORα consensus motif led to an ~30% increase in GATA6 protein levels at the PFG and PP stages of the differentiation when compared to the iPS$^{+/+\,|\,A/A}$ line

(Figures 3-12 B, C, D and Table 3-2). Increased NKX6.1+ cells were observed at the PP stage (45 +/- 6 % vs. 27 +/- 4 %) (Figures 3-12 D, E and Table 3-3). These data confirm that a functioning RORα binding site at rs12953985 in the patient iPSC line enhanced the efficiency of pancreas differentiation to be more comparable to the wild type Mel1 line.

Figure 3-12: Introducing a consensus RORα motif at rs12953985 in the patient iPS cell lines improves differentiation efficiency

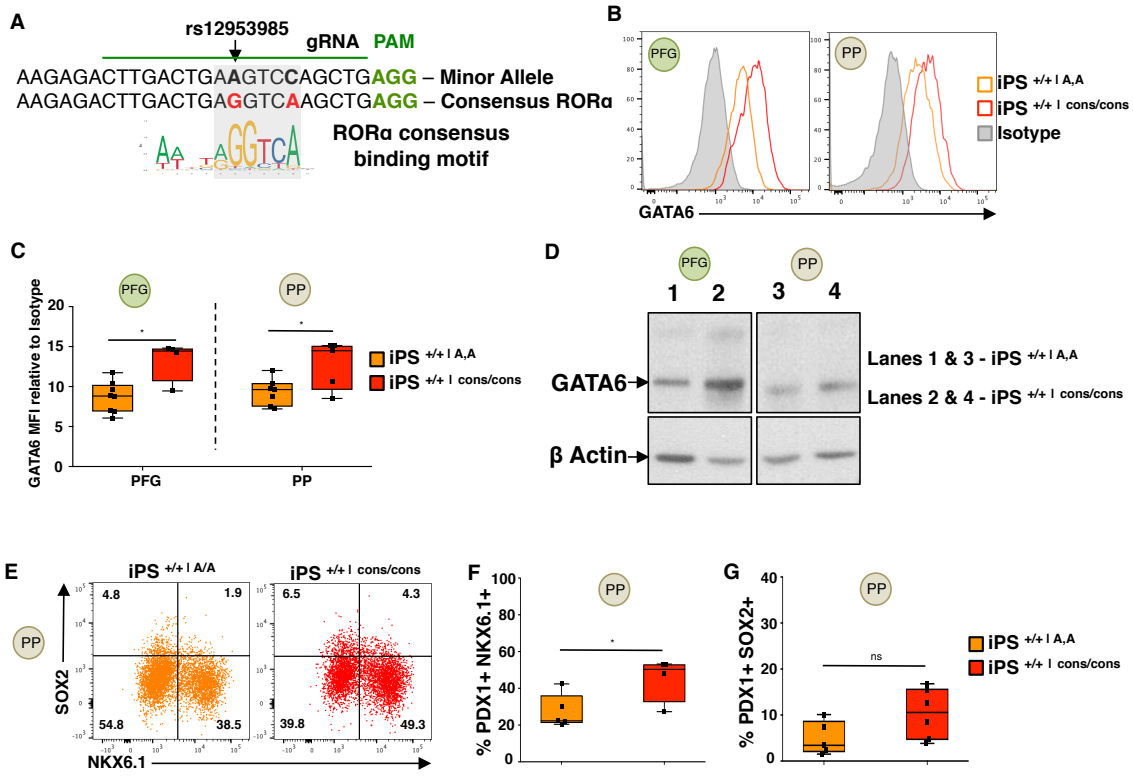
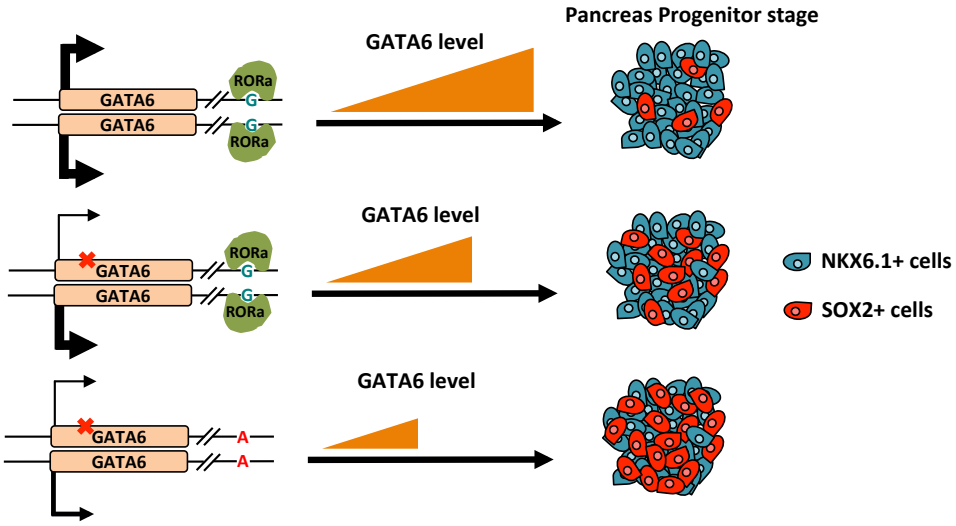


Figure 3-12: Introducing a consensus RORα motif at rs12953985 in the patient iPS cell lines improves differentiation efficiency

(A) Sequence of R5 around rs12953985. Changes to the minor allele variant are highlighted in red. The green line represents the guide RNA used for targeting with the PAM sequence highlighted in green.

B-G) Data from iPS$^{+/+|A/A}$ and iPS$^{+/+|cons/cons}$ lines.
(B) Representative flow cytometry histograms for GATA6 compared to Isotype control at the PFG and PP stages.
(C) Mean fluorescence Intensity (MFI) of GATA6 relative to MFI of Isotype at the PFG and PP stages.
(D) Western blotting for GATA6 and loading control β-Actin at the PFG and PP stages. Lanes 1 and 3 represent iPS$^{+/+|A/A}$ lines at the PFG and PP stages respectively; Lanes 2 and 4 represent iPS$^{+/+|cons/cons}$ lines at the PFG and PP stages respectively.
(E) Representative flow cytometry dot plot for SOX2 and NKX6.1 co-staining at the PP stage.
(F) Quantification of %PDX1+/NKX6.1+ double positive cells at the PP stage.
(G) Quantification of %PDX1+/SOX2+ double positive cells at the PP stage. All data represented as Mean +/- SEM. Ordinary one-way ANOVA with multiple comparisons were used for statistics. *pvalue <0.05, **pvalue <0.01, ***pvalue <0.001, ****pvalue <0.0001.

Figure 3-13: Model of disease mechanism



3.3 Discussion

GWAS have identified many variants located in non-coding genomic regions that are associated with gene expression in specific tissues and diseases (Boyle, Li and Pritchard, 2017). The effect of these cis-acting non-coding variants on gene expression have been proposed as a factor contributing to phenotypic variation of complex traits and disease susceptibility (Castel *et al.*, 2018). In fact, it has been suggested that there could be a large number of disease-associated variants with decreasingly small effects that could act in tandem to contribute to a stronger phenotype (Park *et al.*, 2011). Initiatives such as the GTEx project have identified many functional variants associated with gene expression or eGenes (Consortium *et al.*, 2017). However, these studies depend on primary tissues from post-mortem donors, which are scarce. As a result, information regarding the effects of variants during early human development is not available. Genetically matched isogenic hPSCs lines developed using genome editing

technologies such as CRISPR-CAS9 have proven to be a powerful tool to model the effect of coding mutations during development. This study has expanded this model's applications to include the effects of coding and non-coding variants on gene expression and disease penetrance during development.

Genome edited isogenic human patient-derived iPSCs and ESCs were generated to interrogate the effect of a non-coding SNP on GATA6 expression during pancreas development. This variant, rs12953985, in conjunction with a heterozygous *GATA6* mutation reduced the efficiency of generating pancreatic progenitors *in vitro*. The frequency of the minor allele of the SNP is enriched among patients with PA and was found to be present *in trans* with the mutation in 7 PA patients. Since *GATA6* heterozygous mutations are known to have variable clinical penetrance, the presence of this non-coding variant provides a possible mechanistic explanation for the more severe pancreatic agenesis phenotype, at least in some families. Our cohort mainly included patients with PA (33 versus 10 patients with diabetes or congenital heart disease) caused by *de novo* mutations (the mutation was inherited from a heterozygous parent in 4 cases) (Figure 3-4 and Table 3-4). This means that although our results are consistent with the A allele of the rs12953985 SNP increasing the risk of pancreatic agenesis in patients with pathogenic *GATA6* mutations, further studies in larger cohorts are needed to assess the effect size of the A allele of SNP rs12953985 on the pancreatic agenesis risk and to identify further contributing factors. Other explanations for the varied penetrance of GATA6 heterozygous mutations have been proposed. In the case of de novo mutations, the most likely cause is mosaicism (De Franco, Shaw-Smith, Sarah E. Flanagan, *et al.*, 2013; Yu *et al.*, 2014; Yau *et al.*, 2017). In cases where the mutation is

inherited, dosage levels of GATA4 and/or retinoic acid (RA) have been proposed as possible causes for varied phenotypes in patients (Shi *et al.*, 2017; Tiyaboonchai *et al.*, 2017)

We found that the minor allele variant of SNP rs12953985 leads to the disruption of a RORα binding site. RORα is part of a family of nuclear receptors, also consisting of RORβ and RORγ, that can function as transcription factors to regulate gene expression(Chai *et al.*, 2013). They bind as monomers to the consensus DNA motif RGGTCA and can act as an "orphan" nuclear receptor to activate transcription even in the absence of bound ligands (Chauvet *et al.*, 2011). Ligand binding could lead to the recruitment of either co-activator or co-repressor complexes, which suggests that RORα could act as either a repressor or activator of gene expression (Solt and Burris, 2012). By using a synthetic compound, SR1001, which binds to the ligand-binding domain of RORα and increases the affinity for co-repressors, we were able to show that RORα can regulate GATA6 expression only when then SNP major allele G variant is present during pancreas specification (Figure 3-9 and 3-10)(Solt and Burris, 2012; Lin *et al.*, 2017). However, we do not rule out the possibility that other factors may differentially bind to this site and regulate GATA6 as well. Phenotypes observed in RORα mutant mice revealed a role for RORα in modulating diet-induced obesity, insulin sensitivity and glucose uptake (Lau *et al.*, 2011; Billon, Sitaula and Thomas P. Burris, 2017). Studies have also shown the expression of ROR receptors in pancreatic islet cells and RORα regulates the expression of Ins2 in rat INS-1 cells (Mühlbauer *et al.*, 2013; Kuang *et al.*, 2014). Moreover, RORα was identified as a diabetes susceptibility locus in Mexican Americans and Han Chinese (Hayes *et al.*, 2007; Zhang *et al.*, 2016). Recently, a study

of human embryos from late CS12 to early CS14 using laser capture of the developing pancreas and liver have suggested a specific role for RORα in regulating pancreas development (Jennings *et al.*, 2017). Of the 655 transcription factors identified as key pancreatic regulators in this study, 44% were predicted to be regulated by RORα using motif discovery. These studies highlight the potential role for RORα as a regulator of the pancreatic program in humans. While our study further strengthens these findings, a more detailed analysis is needed to identify direct targets and pathways regulated by RORα during pancreas development to determine if RORα influences other genes regulating pancreas development and function as well as confirm its impacts on GATA6 expression.

Another interesting finding from our studies involves the cell fate switch where the PDX1+ PFG cells develop into SOX2 expressing cells at the expense of NKX6.1 expression. It has been well established that the PDX1+/SOX2+ domain in the developing mouse endoderm at E10.5 gives rise to the antral portion of the stomach (Willet and Mills, 2016; McCracken *et al.*, 2017). Experiments in mouse models of PA caused by knocking out GATA6 and GATA4 have shown that at E10.5 of mouse development, pancreatic lineage cells in the dorsal pancreatic endoderm switch to a stomach identity by expressing SOX2. Similarly, in the ventral pancreatic endoderm, the pancreatic lineages appear to adopt intestinal cell fates by expressing CDX2. This study concludes that the erroneous upregulation of the hedgehog pathway in the GATA6/GATA4 double knock out embryos results in re-specification of the dorsal and ventral pancreatic lineages to the adjacent stomach and intestinal fates, respectively (Xuan and Sussel, 2016). RNAseq data from another group looking at GATA6 mutant

cells during in vitro pancreas differentiation from ESCs also show an upregulation of SOX2 caused by heterozygous GATA6 mutations (Shi *et al.*, 2017). Additionally, in a patient suffering from dorsal PA, contrast CT imaging showed the presence of stomach and bowel loops in the distal pancreatic bed (Sandip *et al.*, 2016). Taken together, these results suggest that GATA6 haploinsufficiency leads to a fate switch in the PDX1+ population from NKX6.1+ pancreas progenitors to SOX2+ antral stomach progenitors.

In summary, we have highlighted the use of patient derived stem cell lines to identify genome variants that may be contributing to changes in gene expression during certain stages of human development. Using genetically identical hPSC lines and CRISPR-CAS9 gene editing, we identified a non-coding SNP that influences GATA6 regulation and could contribute to a stronger PA phenotype. Our results highlight a possible role for RORα during pancreas specification in humans. Further studies will be important to further define the role of the ROR family in pancreas development and identify other non-coding variants in patients that could contribute to disease penetrance.

Table 3-1. hPSC lines generated

Line name	Genetic background	GATA6 allele 1	GATA6 allele 2	rs12953985 Genotype
$Mel^{+/+|G/G}$	Mel1	Wild type	Wild type	G|G
$Mel^{+/mut|G/G}$	Mel1	Wild type	c. 627-630dupCAGT	G|G
$Mel^{+/+|A/A}$	Mel1	Wild type	Wild type	A|A

$Mel^{+/mut|A/A}$	Mel1	Wild type	c. 627-630dupCAGT	A|A
$iPS^{+/+|A/A}$	CHOP.Panagenesis1	Wild type	Wild type	A|A
$iPS^{+/mut|A/A}$	CHOP.Panagenesis1	Wild type	c. 606-609dupGTAC	A|A
$iPS^{+/+|cons/coms}$	CHOP.Panagenesis1	Wild type	Wild type	RORα consensus| RORα consensus
$CHOPWT6^{+/+|G/G}$	CHOPWT6 iPS	Wild type	Wild type	G|G
$CHOPWT6^{+/mut|G/G}$	CHOPWT6 iPS	Wild type	c. 627-630dupCAGT	G|G
$CHOPWT6^{+/+|A/A}$	CHOPWT6 iPS	Wild type	Wild type	A|A
$CHOPWT6^{+/mut|A/A}$	CHOPWT6 iPS	Wild type	c. 627-630dupCAGT	A|A

Table 3-2. GATA6 MFI measurements relative to Isotype in Figures 3-1 G, 3-6 B, 3-9 G and 3-12 C

Line	Stage of Differentiation	GATA6 MFI (Mean +/- SEM)	N (Number of biological replicates)
$iPS^{+/+}$ Figure 3-1 G	DE	11.053 +/- 1.283	5
	PFG	8.664 +/- 0.685	8
	PP	9.330 +/- 0.634	7
$iPS^{+/mut}$ Figure 3-1 G	DE	10.982 +/- 1.183	4
	PFG	6.251 +/- 0.420	10
	PP	7.267 +/- 0.354	7
$Mel^{+/+}$ Figure 3-1 G	DE	11.107 +/- 0.773	6
	PFG	12.335 +/- 1.127	6
	PP	15.385 +/- 0.616	6
$Mel^{+/mut}$ Figure 3-1 G	DE	8.705 +/- 0.922	6
	PFG	8.342 +/- 0.849	6
	PP	9.874 +/- 0.786	7
$Mel^{+/+ \mid G/G}$ Figure 3-6 B	DE	11.107 +/- 0.773	6
	PFG	12.907 +/- 1.031	10
	PP	14.085 +/- 0.828	10

Mel $^{+/+ \mid A/A}$ Figure 3-6 B	DE	9.183 +/- 0.583	5
	PFG	10.046 +/- 0.731	10
	PP	11.439 +/- 0.610	10
Mel $^{+/mut \mid G/G}$ Figure 3-6 B	DE	8.705 +/- 0.922	6
	PFG	8.823 +/- 0.779	9
	PP	9.774 +/- 0.431	6
Mel $^{+/mut \mid A/A}$ Figure 3-6 B	DE	9.471 +/- 0.557	5
	PFG	7.423 +/- 0.514	10
	PP	7.689 +/- 0.452	9
Mel $^{+/+ \mid G/G}$ +DMSO Figure 3-9 G	PFG	12.032 +/- 1.984	4
	PP	10.540 +/- 1.557	4
Mel $^{+/+ \mid G/G}$ +SR1001 Figure 3-9 G	PFG	8.090 +/- 1.340	4
	PP	8.604 +/- 1.798	4
Mel $^{+/+ \mid A/A}$ +DMSO Figure 3-9 G	PFG	6.396+/- 0.754	4
	PP	6.116 +/- 1.858	4
Mel $^{+/+ \mid A/A}$ +SR1001 Figure 3-9 G	PFG	6.166+/- 1.279	4
	PP	6.496 +/- 2.108	4

iPS $^{+/+ \mid A/A}$ Figure 3-12 C	PFG	8.664 +/- 0.685	8
	PP	9.330 +/- 0.634	7
iPS $^{+/+ \mid cons/cons}$ Figure 3-12 C	PFG	13.286 +/- 1.281	4
	PP	12.769 +/- 1.349	5

Table 3-3. Percentage of PDX1+/NKX6.1+ and PDX1+/SOX2+ cells at the PP stage in Figures 3-1 D, 3-1 E, 3-6 E, 3-6 F, 3-9 I, 3-12 F

Line	Type of cells	% Of cells (Mean +/- SEM)	N (Number of biological replicates)
$iPS^{+/+}$ Figures 3-1 D and E	PDX1+ / NKX6.1+	31.08 +/- 3.277	4
	PDX1+ / SOX2+	7.918 +/- 0.8555	5
$iPS^{+/mut}$ Figures 3-1 D and E	PDX1+ / NKX6.1+	10.09 +/- 6.058	4
	PDX1+ / SOX2+	21.38 +/- 3.869	6
$Mel^{+/+}$ Figures 3-1 D and E	PDX1+ / NKX6.1+	60.33 +/- 4.298	7
	PDX1+ / SOX2+	6.800 +/- 1.931	5
$Mel^{+/mut}$ Figures 3-1 D and E	PDX1+ / NKX6.1+	28.15 +/- 2.452	4
	PDX1+ / SOX2+	18.83 +/- 0.9503	8
Mel $^{+/+ \mid G/G}$ Figures 3-6 E and F	PDX1+ / NKX6.1+	60.81 +/- 3.076	14
	PDX1+ / SOX2+	6.800 +/- 1.931	5

Mel $^{+/+ \mid A/A}$ Figures 3-6 E and F	PDX1+ / NKX6.1+	37.77 +/- 3.107	15
	PDX1+ / SOX2+	11.50 +/- 1.439	8
Mel $^{+/mut \mid G/G}$ Figures 3-6 E and F	PDX1+ / NKX6.1+	35.03 +/- 5.371	8
	PDX1+ / SOX2+	18.83 +/- 0.9503	8
Mel $^{+/mut \mid A/A}$ Figures 3-6 E and F	PDX1+ / NKX6.1+	22.47 +/- 4.674	8
	PDX1+ / SOX2+	26.68 +/- 1.972	5
Mel $^{+/+ \mid G/G}$ +DMSO Figure 3-9 I	PDX1+ / NKX6.1+	52.9 +/- 10.18	4
Mel $^{+/+ \mid G/G}$ +SR1001 Figure 3-9 I	PDX1+ / NKX6.1+	35.075 +/- 10.341	4
Mel $^{+/+ \mid A/A}$ +DMSO Figure 3-9 I	PDX1+ / NKX6.1+	39.05 +/- 8.704	4
Mel $^{+/+ \mid A/A}$ +SR1001 Figure 3-9 I	PDX1+ / NKX6.1+	38.55 +/- 9.263	4
iPS $^{+/+ \mid A/A}$ Figure 3-12 E	PDX1+ / NKX6.1+	27.38 +/- 4.111	5
iPS $^{+/+ \mid cons/cons}$ Figure 3-12 E	PDX1+ / NKX6.1+	45.27 +/- 6.116	4

Each biological replicate is a new differentiation from pluripotent stem cells. From every biological replicate, cells were harvested at different stages of the differentiation.

Table 3-4. PA patient information

Patient ID	*GATA6 Mutation*	*De novo*	Pancreatic phenotype	rs12953985 genotype	SNP A allele and mutation *in trans*	Country
Chicago #1	c.1366C>T, p.(Arg456Cys)	Not known (parents not available)	Pancreatic agenesis	A/G	Not known	USA
Chicago #2	c.1087_1097del, p.(Gln363fs)	Not known (parents not available)	Pancreatic agenesis	A/G	Not known	USA
Chicago #3	c.1366C>T, p.(Arg456Cys)	Not known (parents not available)	Pancreatic agenesis	A/A	Yes	USA
Chicago #4	c.1291_1300del, p.(Gln431fs)	yes	Pancreatic agenesis	A/A	Yes	USA
Chicago #5	c.785G>A, p.(Arg262His)	Not known (parents not available)	Diabetes diagnosed at age 23	A/G	Not known	USA
Chicago N6320	c.1366C>T, p.(Arg456Cys)	Not known (parents not available)	Diabetes diagnosed at age 16	A/G	Not known	USA
CHOP. Panagenesis1	c.606_609dup, p.(His204fs)	Not known (parents not available)	Pancreatic agenesis	A/A	Yes	USA

Exeter_01	c.1406G>A, p.(Gly469Glu)	No	Pancreatic agenesis	A/A	Yes	Netherlands
Exeter_01M	NA	NA	NA	A/A	NA	Netherlands
Exeter_01F	c.1406G>A, p.(Gly469Glu)	Not known (parents not available)	Diabetes	G/A	No	Netherlands
Exeter_02	c.1417A>C, p.(Lys473Gln)	Yes	Pancreatic agenesis	G/A	Not known	Netherlands
Exeter_02M	NA	NA	NA	G/G	NA	Netherlands
Exeter_02F	NA	NA	NA	G/A	NA	Netherlands
Exeter_03	c.1303-1G>T, p.?	No	Pancreatic agenesis	G/A	Not known	United Kingdom
Exeter_03M	c.1303-1G>T, p.? Mosaic	Yes	No	G/A	Not known	United Kingdom
Exeter_04	c.635_660del, p.(Pro212fs)	No	Pancreatic agenesis	G/A	Yes	Canada
Exeter_04M	c.635_660del, p.(Pro212fs) Mosaic	Yes	No	G/A	Not known	Canada

Exeter_04F	NA	NA	NA	G/A	NA	Canada
Exeter_04S1	NA	NA	NA	G/A	NA	Canada
Exeter_04S2	c.635_660del, p.(Pro212fs)	No	No (heart disease)	G/G	NA	Canada
Exeter_04S3	c.635_660del, p.(Pro212fs)	No	Small pancreas with normal islets	G/G	NA	Canada
Exeter_05	c.1013C>A, p.(Ser338*)	Yes	Pancreatic agenesis	G/G	NA	United Kingdom
Exeter_05M	NA	NA	NA	G/G	NA	United Kingdom
Exeter_05F	NA	NA	NA	G/G	NA	United Kingdom
Exeter_06	c.(?_-1)_(1788+1_?)del, p.?	Yes	Pancreatic agenesis	A/A	Yes	Germany
Exeter_06M	NA	NA	NA	G/A	NA	Germany
Exeter_06F	NA	NA	NA	G/A	NA	Germany
Exeter_07	c.877_880delinsTAC, p.(Val293fs)	Not known (parents not available)	Pancreatic agenesis	G/A	Not known	USA

Exeter_08	c.1330T>C, p.(Cys444Arg)	No	Pancreatic agenesis	G/G	NA	Australia
Exeter_08M	NA	NA	NA	G/G	NA	Australia
Exeter_08F	c.1330T>C, p.(Cys444Arg)	Not known (parents not available)	No (heart disease)	G/G	NA	Australia
Exeter_09	c.1367G>A, p.(Arg456His)	Yes	Transient neonatal diabetes	G/G	NA	USA
Exeter_09M	NA	NA	NA	G/A	NA	USA
Exeter_09F	NA	NA	NA	G/G	NA	USA
Exeter_10	c.214G>T, p.(Gly72*)	Not known (not in mother, father not available)	Diabetic (diagnosed at 13 years)	G/G	NA	United Kingdom
Exeter_10M	NA	NA	NA	G/G	NA	United Kingdom
Exeter_11	c.1367G>A, p.(Arg456His)	Yes	Pancreatic agenesis	G/G	NA	United Kingdom
Exeter_11M	NA	NA	NA	G/G	NA	United Kingdom

Exeter_11F	NA	NA	NA	G/G	NA	United Kingdom
Exeter_12	c.1396A>G, p.(Asn466Asp)	Yes	Diabetic (diagnosed at 12 years)	A/A	Yes	United Kingdom
Exeter_12M	NA	NA	NA	G/A	NA	United Kingdom
Exeter_12F	NA	NA	NA	A/A	NA	United Kingdom
Exeter_13	c.1108_1121dup, p.(Gly375Sfs)	Yes	Pancreatic agenesis	A/A	Yes	Sweden
Exeter_13M	NA	NA	NA	A/A	NA	Sweden
Exeter_13F	NA	NA	NA	A/A	NA	Sweden
Exeter_14	c.1366C>T, p.(Arg456Cys)	Yes	Pancreatic agenesis	A/A	Yes	United Kingdom
Exeter_14M	NA	NA	NA	G/A	NA	United Kingdom
Exeter_14F	NA	NA	NA	G/A	NA	United Kingdom
Exeter_15	c.1429-41_1441del, p.?	Yes	Pancreatic agenesis	G/G	NA	Germany
Exeter_15M	NA	NA	NA	G/A	NA	Germany

Exeter_15F	NA	NA	NA	G/A	NA	Germany
Exeter_16	c.1429-8T>G, p.?	Yes	Pancreatic agenesis	G/G	NA	Germany
Exeter_16M	NA	NA	NA	G/G	NA	Germany
Exeter_16F	NA	NA	NA	G/G	NA	Germany
Exeter_17	c.1435A>G, p.(Arg479Gly)	Yes	Transient neonatal diabetes	G/G	NA	Netherlands
Exeter_17M	NA	NA	NA	G/A	NA	Netherlands
Exeter_17F	NA	NA	NA	G/G	NA	Netherlands
Exeter_18	c.1498_1501del, p.(Lys500fs)	Yes	Pancreatic agenesis	G/A	Not known	Netherlands
Exeter_18M	NA	NA	NA	G/G	NA	Netherlands
Exeter_18F	NA	NA	NA	A/A	NA	Netherlands
Exeter_19	c.1516+1G>C, p.?	Yes	Pancreatic agenesis	G/A	Not known	USA
Exeter_19M	NA	NA	NA	G/A	NA	USA

Exeter_19F	NA	NA	NA	n/a	NA	USA
Exeter_20	c.1448_1455del, p.(Met483fs)	Yes	Pancreatic agenesis	G/A	Not known	New Zealand
Exeter_20M	NA	NA	NA	G/G	NA	New Zealand
Exeter_20F	NA	NA	NA	G/A	NA	New Zealand
Exeter_21	c.-(265-?_1135+?del), p.?	Not known (not in mother, father not available)	Pancreatic agenesis	G/A	Not known	United Kingdom
Exeter_21M	NA	NA	NA	G/A	NA	United Kingdom
Exeter_22	c.1399G>A, p.(Ala467Thr)	Not known (parents not available)	Pancreatic agenesis	G/G		USA
Exeter_23	c.1036_1042del, p.(Thr346fs)	No	Pancreatic agenesis	A/A	Yes	Germany
Exeter_23M	NA	NA	NA	G/A	NA	Germany
Exeter_23F	c.1036_1042del, p.(Thr346fs)	Not known (not in mother, father not available)	Diabetes (diagnosed at 46)	G/A	No	Germany

Exeter_24	c.701del, p.(Pro234fs)	Yes	Pancreatic agenesis	G/A	Not known	Kosovo
Exeter_24M	NA	NA	NA	G/G	NA	Kosovo
Exeter_24F	NA	NA	NA	G/A	NA	Kosovo
Exeter_25	c.1303-10C>G, p.?	Yes	Pancreatic agenesis	G/G	NA	United Kingdom
Exeter_25M	NA	NA	NA	G/G	NA	United Kingdom
Exeter_25F	NA	NA	NA	G/G	NA	United Kingdom
Exeter_26	c.1396A>G, p.(Asn466Asp)	Not known (not in mother, father not available)	Pancreatic agenesis	G/A	Not known	Australia
Exeter_26M	NA	NA	NA	G/A	NA	Australia
Exeter_27	c.1366C>T, p.(Arg456Cys)	Yes	Pancreatic agenesis	G/G	NA	Canada
Exeter_27M	NA	NA	NA	n/a	NA	Canada
Exeter_27F	NA	NA	NA	n/a	NA	Canada
Exeter_28	c.969C>A, p.(Tyr323	No	Pancreatic agenesis	G/A	Yes	Czech Republic

	*)					
Exeter_28M	NA	NA	NA	A/A	NA	Czech Republic
Exeter_28F	c.969C>A, p.(Tyr323*)	Yes	Diabetes (diagnosed at 30)	G/G	NA	Czech Republic
Exeter_29	c.1296del, p.(Lys432fs)	Yes	Pancreatic agenesis	G/A	Not known	United Kingdom
Exeter_29M	NA	NA	NA	G/G	NA	United Kingdom
Exeter_29F	NA	NA	NA	A/A	NA	United Kingdom
Exeter_30	c.744del, p.(Pro249fs)	Yes	Pancreatic agenesis	G/A	Not known	Canada
Exeter_30M	NA	NA	NA	G/A	NA	Canada
Exeter_30F	NA	NA	NA	G/G	NA	Canada
Exeter_31	c.1369A>G, p.(Arg457Gly)	Yes	Pancreatic agenesis	G/A	Not known	USA
Exeter_31M	NA	NA	NA	G/G	NA	USA
Exeter_31F	NA	NA	NA	G/A	NA	USA

CHAPTER 4: SUMMARY AND FUTURE DIRECTIONS

4.1 Other non-coding disease modifiers of PA

The major finding in this study is the disease modifying SNP rs12953985. The minor allele variant of this SNP in conjunction with a heterozygous mutation in the GATA6 coding region leads to a stronger PA disease phenotype. In this study we found that the minor allele, A, was more frequent among patients with PA compared to the non-PA group. The A allele was present in 25/33 individuals with *GATA6* PA tested in the two cohorts for a cumulative allele frequency of 48.5%. This is higher than the allele frequency among patients with the *GATA6* mutation without PA (27.8%) and the frequency among Europeans in the GnomAD database (33.5%).

We also found cases where patients suffer from PA while having the major allele variant, G, of rs12953985. We hypothesize that in these cases there may be other direct genetic modifiers of GATA6 expression or a modifier of genes upstream of GATA6. If this is the case, a more detailed study of PA patient genomes using whole genome sequencing or deep sequencing of regions of interest around pancreas specific genes would be needed to identify other modifiers. In this study, we have identified 6 non-coding regions of interest based on histone marks and transcription factor binding. While we did not find any variants in this region in our CHOP.Panagenesis iPS cell line, these regions could hold variants of interest in other patient samples. Sequencing these regions will yield a deeper understanding on the different variants that could contribute to the more severe PA phenotype.

Previous studies have also shown that retinoic acid (RA) signaling and GATA4 expression levels could play a role in the severity of phenotype (Shi *et al.*, 2017; Tiyaboonchai *et al.*, 2017). In the case of morphogen signaling pathways like RA, lower doses of RA in hPSC differentiation models led to lower GATA6 levels, which in turn led to fewer pancreas progenitors. It would be interesting to screen for variants in RA pathway genes in our patient cohort. Additionally, this study and others have shown that GATA factors regulate each other; specifically GATA6 regulates GATA4 during human pancreas development. In some cases, an additional variant in the GATA4 locus could be leading to the more severe PA phenotype. While we have identified one disease modifying variant of PA, we believe further research will yield more such modifiers and provide a deeper understanding of the pathogenesis of PA.

4.2 Identifying additional downstream targets of GATA6

While this study has highlighted an upstream regulator of GATA6 expression, some questions regarding the downstream targets of GATA6 are still unexplored. We have shown that mutations in GATA6 lead to a decrease in GATA6 expression levels, which in turn cause a reduction in pancreas progenitor gene signature, specifically in genes such as GATA4, PDX1, NKX6-1 and others. We have also shown that this causes an increase in stomach progenitor gene signature, specifically in SOX2 and IRX2.

What remains to be understood is how GATA6 regulates the expression of these genes. This could be either by direct regulation of the transcription of these genes or an indirect regulation or a combination of the two. These questions can be best answered by performing ChIP-seq for GATA6 at the foregut stage of the differentiation. This would

help elucidate the potential direct targets of GATA6. Once we identify the regions that show GATA6 binding, we can then measure RNA and protein levels of nearby genes to conclusively answer the question of direct vs indirect regulation. If this approach is inconclusive, an unbiased approach of an RNA-seq could be very useful to elucidate all the genes dysregulated by GATA6 haploinsufficiency. We could then use gene ontology to identify potential gene networks and signaling pathways that may be affected by GATA6 mutations. Using this information, we can tweak these signaling pathways using small molecules during the hPSC pancreas differentiation and attempt to rescue the pancreas progenitor phenotype.

Additionally, if we identify GATA6 binding in the regulatory regions of SOX2 during pancreas development, this could suggest a direct regulation of SOX2 by GATA6, providing a mechanism for the fate switch from pancreas to stomach progenitors. To test whether GATA6 is repressing SOX2 expression by binding at regulatory regions, we can modify these GATA6 binding motifs and observe the effect on SOX2 expression levels. Loss of GATA6 binding to these mutated regions should lead to an increase in SOX2 gene expression specifically at the pancreas progenitor stage of the differentiation.

These studies could also provide some insight into the differences in gene regulation between humans and murine models. As pancreas differentiation protocols are becoming more sophisticated and we are getting closer to generating a functional adult β-cells from PSCs, any deeper insight into human specific pancreas gene regulation will be invaluable. This study has highlighted a key difference in human and mice pancreas development at a human specific regulatory region. We may also be able to decipher

human specific gene regulatory networks governed by GATA6 that leads to a more severe phenotype in humans as compared to mice.

4.3 Elucidating the role of GATA6 in gut tube patterning

One of the most interesting findings in this study is the switch in the cell fate of PP cells with lower GATA6 levels. We have shown that with decreasing GATA6 levels the PDX1+ PFG cells develop into SOX2 expressing cells at the expense of NKX6.1 expression. Our data supports the idea that GATA6 is involved in anterior/posterior patterning of the gut tube, which is consistent with our previous work looking at GATA6 in anterior endoderm patterning (Liao *et al.*, 2018). In both cases decreased expression of GATA6 leads to marker expression indicative of anteriorization (Figure 4-1).

A further mechanistic analysis of how GATA6 patterns the gut tube is the logical next step. In mice it has been shown that GATA4 and GATA6 are essential for maintaining repression of hedgehog signaling in the gut tube (Xuan and Sussel, 2016). GATA4 and GATA6 are able to repress transcription through the sonic hedgehog (*Shh*) endoderm-specific enhancer *MACS1* and that GATA-binding sites within this enhancer are necessary for this repressive activity. This leads to a switch in cell fate of pancreas progenitor cells in the developing mice gut tube. To confirm this in the hPSC model system, we would need to increase the external *Shh* inhibition with lowering doses of GATA6 in order to rescue the phenotype. Conversely, removing the exogenous *Shh* inhibition under wild type conditions with normal levels of GATA6 should lead to no change in pancreas specification.

Figure 4-1: Elucidating the role of GATA6 in gut tube patterning

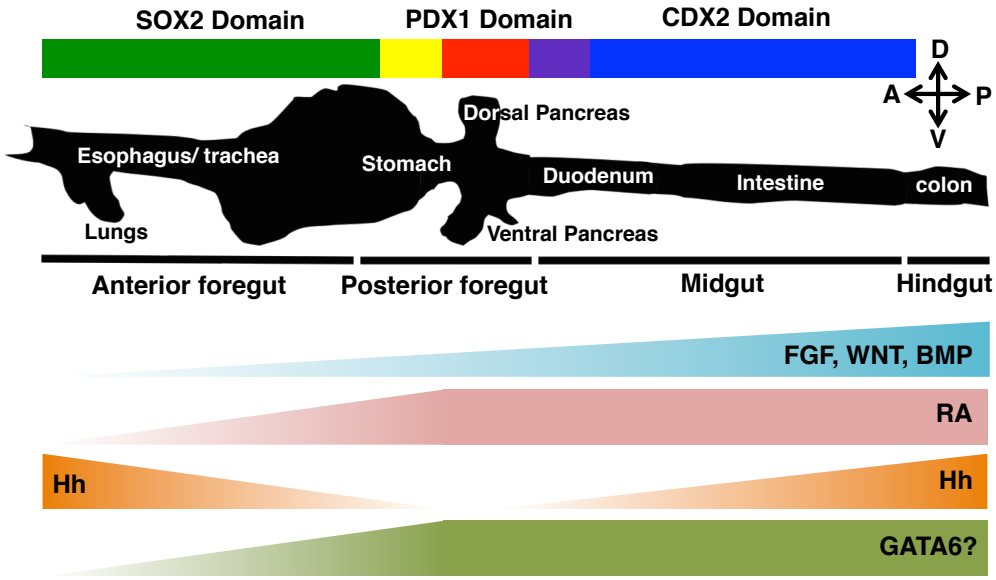


Figure 4-1: Elucidating the role of GATA6 in gut tube patterning

The foregut, midgut and hindgut segments of the gut tube are indicated along with the budding organs from each of these patterned segments. Morphogen gradients of FGF, Wnt, BMP, RA and Hedgehog (Hh) play a key role in specifying where these organ buds develop. These are marked by transcription factors such as SOX2, PDX1 and CDX2. *GATA6* mutations, which lead to lower levels of GATA6 protein, cause an anteriorization phenotype, which suggests that GATA6 acts as a posteriorizing factor during gut tube patterning. The exact mechanism of this is still unknown and future work will determine what morphogen gradients are regulated by GATA6.

Another possible explanation for the patterning defect caused by GATA6 could by dysregulation of the Wnt signaling pathway. RNAseq data from the PP stage in GATA6 deficient cells show a number of Wnt pathway genes that are dysregulated (Shi *et al.*, 2017). These include upregulation of different Wnts and their receptors. Additionally in the mouse embryos, loss of Wnt/β-catenin signaling in the developing stomach leads to posteriorization of the fundus to the antrum (McCracken *et al.*, 2017). β-catenin activation in hPSC-derived foregut progenitors promoted the anteriorization of

developing stomach and the development of human fundic-type gastric organoids (hFGOs). To test whether lower GATA6 levels lead to an upregulation of Wnt/β-catenin signaling pathway, genes that are downstream of this pathway need to be measured at the PFG and PP stages of the pancreas differentiation. Additionally, the anteriorization phenotype that we observe should be rescued by using a Wnt inhibitor at these specific stages.

These studies would provide important implications for the pathways regulated by the GATA factors during human gut tube development and could lead to novel strategies to detect and/or prevent PA before birth.

4.4 Deciphering the role of GATA6 in stomach development

We have identified an increase in antral stomach markers SOX2 and IRX2 in the pancreas progenitor cells with lower GATA6 levels. This finding brings into question the role of GATA6 during human stomach development. Experiments in mouse models of PA caused by knocking out GATA6 and GATA4 have shown that at E10.5 of mouse development, pancreatic lineage cells in the dorsal pancreatic endoderm switch to a stomach identity by expressing SOX2 (Xuan *et al.*, 2012). RNAseq data from another group looking at GATA6 mutant cells during in vitro pancreas differentiation from ESCs also show an upregulation of SOX2 caused by heterozygous GATA6 mutations (Shi *et al.*, 2017). Additionally, in a patient suffering from dorsal PA, contrast CT imaging showed the presence of stomach and bowel loops in the distal pancreatic bed (Sandip *et al.*, 2016).

The role of GATA factors in human stomach development is not very well understood. Stomach differentiation protocols from hPSCs to study stomach development have been a fairly recent advancement in the field. Our genetically matched cell lines with decreasing GATA6 levels may prove to be a perfect model system to understand human stomach development. The hypothesis is that with lower levels of GATA6 we see more of a bias towards fundic cell fate as opposed to an antral cell fate. There are a few experiments of interest. Firstly, it would be interesting to see if the regulatory regions that we identified in our study are also active during stomach development. Secondly, the stage specific regulation of stomach development by GATA6 would give us insights into time points at which the GATA factors are necessary.

GATA6 and GATA4 have been shown to play a role in gastric cancers and more recently there is burgeoning evidence indicating their importance in stomach development. The genetic models that we have developed in this study can be very useful to develop a deeper understanding of the role of GATA factors during human stomach development.

4.5 Understanding the role of RORα during human pancreas development

Our study highlights the potential role for RORα as a regulator of the pancreatic program in humans. We have shown that RORα regulates GATA6 expression specifically during pancreas development by binding to a downstream regulatory region. While this is a good initial step into dissecting the role of RORα, a deeper dive into the function of RORα during human pancreas development is important.

To this date, the most compelling evidence for the specific role of RORα in human pancreas development is from a recent study of human embryos from late CS12 to early CS14 using laser capture of the developing pancreas and liver (Jennings *et al.*, 2017). Of the 655 transcription factors identified as key pancreatic regulators in this study, 44% were predicted to be regulated by RORα using motif discovery. This study uses a purely bioinformatics approach to decipher the potential role of RORα in human pancreas development. To achieve a more detailed understanding, using genome-edited hPSCs with genetically matched RORα mutant and wild type lines will be beneficial. One particular downside of knocking out RORα is that other members of the ROR family, which bind to a highly conserved DNA binding motif, could compensate its role. If this were to be the case, knocking our RORα will have little to no phenotype on pancreas development. To circumvent this, generating a specific dominant negative form of RORα may be beneficial. This would have a similar effect to the inverse agonist SR1001 that we used in our study. A genetically modified cell line will also have fewer off-target effects compared to a small molecule like SR1001. Additionally, to identify specific targets, ChIP-seq for RORα at the FG and PP stages of the differentiation should be performed. These binding data would also need to be correlated with gene expression studies to identify gene networks that are controlled by RORα.

While our study further strengthens the role of RORα during human pancreas development, a more detailed analysis is needed to identify direct targets and pathways regulated by RORα and to determine if RORα influences other genes regulating pancreas development and function as well as confirm its impacts on GATA6 expression.

4.6 Involvement of other nuclear receptors in regulating GATA6

We have identified an orphan nuclear receptor, RORα, which regulates GATA6 by binding to a downstream regulatory region. RORα has previously been implicated in insulin sensitivity, diet-induced obesity and glucose uptake in mice (Lau *et al.*, 2011; Billon, Sitaula and Thomas P. Burris, 2017). Interestingly RORα has also been identified as a diabetes susceptibility locus in Mexican Americans and Han Chinese (Hayes *et al.*, 2007; Zhang *et al.*, 2016). While there have been links between RORα and pancreatic diseases, the phenotypes are not as severe as those seen with GATA6 mutations. One reason for this may be the redundancy between nuclear receptors and the fact that other nuclear receptors bind to similar DNA binding motifs and may compensate for the loss of RORα.

Recent studies have shown that other nuclear receptors can also bind as monomers to the RGGTCA motif (Quack and Carlberg, 2001; Penvose *et al.*, 2019). Of particular interest is another orphan nuclear receptor, NR5A2, which has been shown to play a role in pancreas development and the maintenance of pancreatic exocrine identity (Hale *et al.*, 2014). NR5A2 is a member of the nuclear hormone receptor family and they bind as monomers to extended half-site DNA response elements. NR5A2 has been shown to be crucial during fetal organogenesis for the formation of the pancreas. It controls the expansion of the nascent pancreatic epithelium, the proper formation of the pancreas progenitor population that gives rise to pre-acinar cells and bipotent cells with ductal and islet endocrine potential, and the proper formation and differentiation of pre-acinar cells. NR5A2 has also been shown to play a critical role in regulating pancreatic transcription factors such as GATA4 and FOXA2 (Hale *et al.*, 2014). NR5A2 may be playing a

compensatory role in regulating GATA6 in the cases where mutations lead to loss of RORα function.

To test this hypothesis, we could generate a RORα knockout in the context of wild type and GATA6 heterozygous mutations. If these lines do not have the same phenotype as the minor allele of SNP rs12953985, we could look for compensation by binding of NR5A2 to this site.

BIBLIOGRAPHY

Ahonen, T. *et al.* (2009) 'Gender differences relating to metabolic syndrome and proinflammation in Finnish subjects with elevated blood pressure', *Mediators of inflammation*. 2009/08/20. Hindawi Publishing Corporation, 2009, p. 959281. doi: 10.1155/2009/959281.

Ali, O. (2013) 'Genetics of type 2 diabetes.', *World journal of diabetes*. United States, 4(4), pp. 114–123. doi: 10.4239/wjd.v4.i4.114.

Anik, A. *et al.* (2015) 'Maturity-onset diabetes of the young (MODY): an update.', *Journal of pediatric endocrinology & metabolism : JPEM*. Germany, 28(3–4), pp. 251–263. doi: 10.1515/jpem-2014-0384.

Apelqvist, A., Ahlgren, U. and Edlund, H. (1997) 'Sonic hedgehog directs specialised mesoderm differentiation in the intestine and pancreas.', *Current biology : CB*. England, 7(10), pp. 801–804. doi: 10.1016/s0960-9822(06)00340-x.

Ashcroft, F. M. and Rorsman, P. (2012) 'Diabetes mellitus and the beta cell: the last ten years.', *Cell*. United States, 148(6), pp. 1160–1171. doi: 10.1016/j.cell.2012.02.010.

Association, A. D. (2018) 'Economic Costs of Diabetes in the U.S. in 2017', *Diabetes care*. 2018/03/22. American Diabetes Association, 41(5), pp. 917–928. doi: 10.2337/dci18-0007.

Atkinson, M. A. (2012) 'The pathogenesis and natural history of type 1 diabetes.', *Cold Spring Harbor perspectives in medicine*. United States, 2(11). doi: 10.1101/cshperspect.a007641.

Avior, Y., Sagi, I. and Benvenisty, N. (2016) 'Pluripotent stem cells in disease modelling and drug discovery.', *Nature reviews. Molecular cell biology*. England, 17(3), pp. 170–182. doi: 10.1038/nrm.2015.27.

Ballian, N. and Brunicardi, F. C. (2007) 'Islet vasculature as a regulator of endocrine pancreas function.', *World journal of surgery*. United States, 31(4), pp. 705–714. doi: 10.1007/s00268-006-0719-8.

van Belle, T. L., Coppieters, K. T. and von Herrath, M. G. (2011) 'Type 1 diabetes: etiology, immunology, and therapeutic strategies.', *Physiological reviews*. United States, 91(1), pp. 79–118. doi: 10.1152/physrev.00003.2010.

Bernardo, A. S. *et al.* (2009) 'Biphasic induction of Pdx1 in mouse and human embryonic stem cells can mimic development of pancreatic beta-cells.', *Stem cells (Dayton, Ohio)*. United States, 27(2), pp. 341–351. doi: 10.1634/stemcells.2008-0310.

Biel, N. M. *et al.* (2015) 'Vascular Smooth Muscle Cells From Hypertensive Patient-Derived Induced Pluripotent Stem Cells to Advance Hypertension Pharmacogenomics', *Stem cells translational medicine*. 2015/10/22. AlphaMed Press, 4(12), pp. 1380–1390. doi: 10.5966/sctm.2015-0126.

Billon, C., Sitaula, S. and Burris, Thomas P (2017) 'Metabolic Characterization of a Novel

RORalpha Knockout Mouse Model without Ataxia.', *Frontiers in endocrinology*. Switzerland, 8, p. 141. doi: 10.3389/fendo.2017.00141.

Billon, C., Sitaula, S. and Burris, Thomas P. (2017) 'Metabolic characterization of a novel RORα knockout mouse model without ataxia', *Frontiers in Endocrinology*, 8, p. 141. doi: 10.3389/fendo.2017.00141.

Bonal, C. and Herrera, P. L. (2008) 'Genes controlling pancreas ontogeny.', *The International journal of developmental biology*. Spain, 52(7), pp. 823–835. doi: 10.1387/ijdb.072444cb.

Bonnefond, A. *et al.* (2012) 'GATA6 inactivating mutations are associated with heart defects and, inconsistently, with pancreatic agenesis and diabetes.', *Diabetologia*. Germany, pp. 2845–2847. doi: 10.1007/s00125-012-2645-7.

Boyle, E. A., Li, Y. I. and Pritchard, J. K. (2017) 'An Expanded View of Complex Traits: From Polygenic to Omnigenic.', *Cell*. United States, 169(7), pp. 1177–1186. doi: 10.1016/j.cell.2017.05.038.

Brewer, a *et al.* (1999) 'The human and mouse GATA-6 genes utilize two promoters and two initiation codons.', *The Journal of biological chemistry*, 274(53), pp. 38004–16. doi: 10.1074/jbc.274.53.38004.

Brewer, M. H. *et al.* (2014) 'Haplotype-specific modulation of a SOX10 / CREB response element at the Charcot – Marie – Tooth disease type 4C locus SH3TC2', 23(19), pp. 5171–5187. doi: 10.1093/hmg/ddu240.

Brissova, M., Fowler, M. J., *et al.* (2005) 'Assessment of human pancreatic islet architecture and composition by laser scanning confocal microscopy.', *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. United States, 53(9), pp. 1087–1097. doi: 10.1369/jhc.5C6684.2005.

Brissova, M., Blaha, M., *et al.* (2005) 'Reduced PDX-1 expression impairs islet response to insulin resistance and worsens glucose homeostasis.', *American journal of physiology. Endocrinology and metabolism*. United States, 288(4), pp. E707-14. doi: 10.1152/ajpendo.00252.2004.

Cabrera, O. *et al.* (2006) 'The unique cytoarchitecture of human pancreatic islets has implications for islet cell function.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 103(7), pp. 2334–2339. doi: 10.1073/pnas.0510790103.

Carrasco, M. *et al.* (2012) 'GATA4 and GATA6 control mouse pancreas organogenesis', *Journal of Clinical Investigation*, 122(10), pp. 3504–3515. doi: 10.1172/JCI63240DS1.

Castel, S. E. *et al.* (2018) 'Modified penetrance of coding variants by cis-regulatory variation contributes to disease risk', *Nature Genetics*, 50(9), pp. 1327–1334. doi: 10.1038/s41588-018-0192-y.

Catli, G. *et al.* (2013) 'A novel GATA6 mutation leading to congenital heart defects and permanent neonatal diabetes: a case report.', *Diabetes & metabolism*. Elsevier Masson SAS, 39(4), pp. 370–4. doi: 10.1016/j.diabet.2013.01.005.

Cebola, I. *et al.* (2015) 'TEAD and YAP regulate the enhancer network of human embryonic pancreatic progenitors.', *Nature cell biology*, 17(5), pp. 615–26. doi: 10.1038/ncb3160.

Chai, W. *et al.* (2013) 'The pathophysiological roles and regulatory mechanism of retinoic acid receptor-related orphan receptors.', *Sheng li ke xue jin zhan [Progress in physiology]*. China, 44(2), pp. 87–92.

Chan, J. C. N. and Luk, A. O. Y. (2016) 'Diabetes: A Cinderella Subject We Can't Afford to Ignore.', *PLoS medicine*. United States, 13(7), p. e1002068. doi: 10.1371/journal.pmed.1002068.

Chao, Christina S *et al.* (2015) 'Novel GATA6 mutations in patients with pancreatic agenesis and congenital heart malformations.', *PloS one*. United States, 10(2), p. e0118449. doi: 10.1371/journal.pone.0118449.

Chao, Christina S. *et al.* (2015) 'Novel GATA6 Mutations in Patients with Pancreatic Agenesis and Congenital Heart Malformations', *Plos One*, 10(2), p. e0118449. doi: 10.1371/journal.pone.0118449.

Chauvet, C. *et al.* (2011) 'Control of gene expression by the retinoic Acid-Related Orphan Receptor alpha in HepG2 human hepatoma cells', *PLoS ONE*, 6(7). doi: 10.1371/journal.pone.0022545.

Chen, X. F. *et al.* (2018) 'An Osteoporosis Risk SNP at 1p36.12 Acts as an Allele-Specific Enhancer to Modulate LINC00339 Expression via Long-Range Loop Formation', *American Journal of Human Genetics*. ElsevierCompany., 102(5), pp. 776–793. doi: 10.1016/j.ajhg.2018.03.001.

Collombat, P. *et al.* (2005) 'The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the α- and β-cell lineages in the mouse endocrine pancreas', *Development*, 132(13), pp. 2969 LP – 2980. doi: 10.1242/dev.01870.

Concannon, P., Rich, S. S. and Nepom, G. T. (2009) 'Genetics of type 1A diabetes.', *The New England journal of medicine*. United States, 360(16), pp. 1646–1654. doi: 10.1056/NEJMra0808284.

Concepcion, J. P. *et al.* (2014) 'Neonatal diabetes, gallbladder agenesis, duodenal atresia, and intestinal malrotation caused by a novel homozygous mutation in RFX6.', *Pediatric diabetes*. Denmark, 15(1), pp. 67–72. doi: 10.1111/pedi.12063.

Conrad, E., Stein, R. and Hunter, C. S. (2014) 'Revealing transcription factors during human pancreatic beta cell development.', *Trends in endocrinology and metabolism: TEM*. United States, 25(8), pp. 407–414. doi: 10.1016/j.tem.2014.03.013.

Consortium, Gte. *et al.* (2017) 'Genetic effects on gene expression across human tissues', *Nature*. The Author(s), 550, p. 204. Available at: https://doi.org/10.1038/nature24277.

Consortium, R. E. *et al.* (2015) 'Integrative analysis of 111 reference human epigenomes', *Nature*. The Author(s), 518, p. 317. Available at: https://doi.org/10.1038/nature14248.

Cooper, D. N. *et al.* (2013) 'Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease.', *Human genetics*. Germany, 132(10), pp. 1077–1130. doi: 10.1007/s00439-013-1331-2.

Coppieters, K. T. *et al.* (2012) 'Immunology in the clinic review series: focus on type 1 diabetes and viruses: the role of viruses in type 1 diabetes: a difficult dilemma.', *Clinical and experimental immunology*. England, 168(1), pp. 5–11. doi: 10.1111/j.1365-2249.2011.04554.x.

D'Amour, K. A. *et al.* (2005) 'Efficient differentiation of human embryonic stem cells to definitive endoderm.', *Nature biotechnology*. United States, 23(12), pp. 1534–1541. doi: 10.1038/nbt1163.

D'Amour, K. A. *et al.* (2006) 'Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells.', *Nature biotechnology*. United States, 24(11), pp. 1392–1401. doi: 10.1038/nbt1259.

Decker, K. *et al.* (2006) 'Gata6 is an important regulator of mouse pancreas development.', *Developmental biology*. United States, 298(2), pp. 415–429. doi: 10.1016/j.ydbio.2006.06.046.

DeFronzo, R. A. *et al.* (2015) 'Type 2 diabetes mellitus.', *Nature reviews. Disease primers*. England, 1, p. 15019. doi: 10.1038/nrdp.2015.19.

Degner, J. F. *et al.* (2012) 'DNase I sensitivity QTLs are a major determinant of human expression variation.', *Nature*. England, 482(7385), pp. 390–394. doi: 10.1038/nature10808.

Diaz-Valencia, P. A., Bougneres, P. and Valleron, A.-J. (2015) 'Global epidemiology of type 1 diabetes in young adults and adults: a systematic review.', *BMC public health*. England, 15, p. 255. doi: 10.1186/s12889-015-1591-y.

Do, C. *et al.* (2017) 'Genetic – epigenetic interactions in cis : a major focus in the post-GWAS era', *Genome biology*. Genome Biology, 18(1), p. 120. doi: 10.1186/s13059-017-1250-y.

Dolensek, J., Rupnik, M. S. and Stozer, A. (2015) 'Structural similarities and differences between the human and the mouse pancreas.', *Islets*. United States, 7(1), p. e1024405. doi: 10.1080/19382014.2015.1024405.

Dong, P. D. S. *et al.* (2008) 'Graded levels of Ptf1a differentially regulate endocrine and exocrine fates in the developing pancreas.', *Genes & development*. United States, 22(11), pp. 1445–1450. doi: 10.1101/gad.1663208.

Donnelly, C. J. *et al.* (2013) 'RNA toxicity from the ALS/FTD C9ORF72 expansion is mitigated by antisense intervention.', *Neuron*. United States, 80(2), pp. 415–428. doi: 10.1016/j.neuron.2013.10.015.

Dourov, N. and Buyl-Strouvens, M. L. (1969) '[Agenesia of the pancreas. Anatomo-clinical observations of a case of diabetes mellitus, with steatorrhea and hypotrophy, in a newborn infant].', *Archives francaises de pediatrie*. France, 26(6), pp. 641–650.

Dufort, D. *et al.* (1998) 'The transcription factor HNF3beta is required in visceral endoderm for normal primitive streak morphogenesis.', *Development (Cambridge, England)*. England, 125(16),

pp. 3015–3025.

Ebert, A. D. *et al.* (2014) 'Characterization of the molecular mechanisms underlying increased ischemic damage in the aldehyde dehydrogenase 2 genetic polymorphism using a human induced pluripotent stem cell model system.', *Science translational medicine*. United States, 6(255), p. 255ra130. doi: 10.1126/scitranslmed.3009027.

Eknoyan, G. and Nagy, J. (2005) 'A history of diabetes mellitus or how a disease of the kidneys evolved into a kidney disease.', *Advances in chronic kidney disease*. United States, 12(2), pp. 223–229. doi: 10.1053/j.ackd.2005.01.002.

Ellard, S. *et al.* (2013) 'Improved genetic testing for monogenic diabetes using targeted next-generation sequencing.', *Diabetologia*. Germany, 56(9), pp. 1958–1963. doi: 10.1007/s00125-013-2962-5.

Emison, E. S. *et al.* (2005) 'A common sex-dependent mutation in a RET enhancer underlies Hirschsprung disease risk.', *Nature*. England, 434(7035), pp. 857–863. doi: 10.1038/nature03467.

Fajans, S. S., Bell, G. I. and Polonsky, K. S. (2001) 'Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young.', *The New England journal of medicine*. United States, 345(13), pp. 971–980. doi: 10.1056/NEJMra002168.

Fang, L. *et al.* (2019) 'Functional Genomics and CRISPR Applied to Cardiovascular Research and Medicine', *Arteriosclerosis, Thrombosis, and Vascular Biology*. American Heart Association, 39(9), pp. e188–e194. doi: 10.1161/ATVBAHA.119.312579.

De Franco, E., Shaw-Smith, C., Flanagan, Sarah E, *et al.* (2013) 'GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency.', *Diabetes*, 62(3), pp. 993–7. doi: 10.2337/db12-0885.

De Franco, E., Shaw-Smith, C., Flanagan, Sarah E., *et al.* (2013) 'GATA6 mutations cause a broad phenotypic spectrum of diabetes from pancreatic agenesis to adult-onset diabetes without exocrine insufficiency', *Diabetes*, 62(3), pp. 993–997. doi: 10.2337/db12-0885.

Fuchsberger, C. *et al.* (2016) 'The genetic architecture of type 2 diabetes.', *Nature*. England, 536(7614), pp. 41–47. doi: 10.1038/nature18642.

Gabbay, M. *et al.* (2017) 'Pancreatic Agenesis due to Compound Heterozygosity for a Novel Enhancer and Truncating Mutation in the PTF1A Gene', *Journal of clinical research in pediatric endocrinology*, 9(3), pp. 274–277.

Gadue, P. *et al.* (2006) 'Wnt and TGF-beta signaling are required for the induction of an in vitro model of primitive streak formation using embryonic stem cells.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 103(45), pp. 16806–16811. doi: 10.1073/pnas.0603916103.

Gasa, R. *et al.* (2004) 'Proendocrine genes coordinate the pancreatic islet differentiation program in vitro', *Proceedings of the National Academy of Sciences of the*

United States of America, 101(36), pp. 13245 LP – 13250. doi: 10.1073/pnas.0405301101.

Girlovanu, M. *et al.* (2015) 'Stem cells - biological update and cell therapy progress', *Clujul medical (1957)*. 2015/07/01. Iuliu Hatieganu University of Medicine and Pharmacy, 88(3), pp. 265–271. doi: 10.15386/cjmed-483.

Gittes, G. K. (2009) 'Developmental biology of the pancreas: a comprehensive review.', *Developmental biology*. United States, 326(1), pp. 4–35. doi: 10.1016/j.ydbio.2008.10.024.

Gordon, C. T. and Lyonnet, S. (2014) 'Enhancer mutations and phenotype modularity', *Nature Publishing Group*. Nature Publishing Group, 46(1), pp. 3–4. doi: 10.1038/ng.2861.

Greeley, S. A. W. *et al.* (2010) 'Update in neonatal diabetes.', *Current opinion in endocrinology, diabetes, and obesity*. England, 17(1), pp. 13–19. doi: 10.1097/MED.0b013e328334f158.

Gregg, T. *et al.* (2016) 'Pancreatic beta-Cells From Mice Offset Age-Associated Mitochondrial Deficiency With Reduced KATP Channel Activity.', *Diabetes*. United States, 65(9), pp. 2700–2710. doi: 10.2337/db16-0432.

Gurdon, J. B. (1962) 'The Developmental Capacity of Nuclei taken from Intestinal Epithelium Cells of Feeding Tadpoles', *Journal of Embryology and Experimental Morphology*, 10(4), pp. 622 LP – 640. Available at: http://dev.biologists.org/content/10/4/622.abstract.

Guz, Y. *et al.* (1995) 'Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny.', *Development (Cambridge, England)*. England, 121(1), pp. 11–18.

Habener, J. F., Kemp, D. M. and Thomas, M. K. (2005) 'Minireview: Transcriptional Regulation in Pancreatic Development', *Endocrinology*, 146(3), pp. 1025–1034. doi: 10.1210/en.2004-1576.

Hale, M. A. *et al.* (2014) 'The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis', *Development*, 141(16), pp. 3123 LP – 3133. doi: 10.1242/dev.109405.

Hamazaki, T. *et al.* (2017) 'Concise Review: Induced Pluripotent Stem Cell Research in the Era of Precision Medicine.', *Stem cells (Dayton, Ohio)*. United States, 35(3), pp. 545–550. doi: 10.1002/stem.2570.

Harries, L. W. *et al.* (2006) 'Isomers of the TCF1 gene encoding hepatocyte nuclear factor-1 alpha show differential expression in the pancreas and define the relationship between mutation position and clinical phenotype in monogenic diabetes.', *Human molecular genetics*. England, 15(14), pp. 2216–2224. doi: 10.1093/hmg/ddl147.

Hattersley, A. T. and Patel, K. A. (2017) 'Precision diabetes: learning from monogenic diabetes.', *Diabetologia*. Germany, 60(5), pp. 769–777. doi: 10.1007/s00125-017-4226-2.

Haumaitre, C. *et al.* (2005) 'Lack of TCF2/vHNF1 in mice leads to pancreas agenesis.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 102(5), pp. 1490–1495. doi: 10.1073/pnas.0405776102.

Hayes, M. G. *et al.* (2007) 'Identification of Type 2 Diabetes Genes in Mexican Americans Through Genome-Wide Association Studies', *Diabetes*, 56(12), pp. 3033 LP – 3044. doi: 10.2337/db07-0482.

Hebrok, M., Kim, S. and Jacques, B. S. (2000) 'Regulation of pancreas development by hedgehog signaling', *Development*, 127, pp. 4905–4913. Available at: http://dev.biologists.org/content/127/22/4905.short.

Heintzman, N. D. *et al.* (2007) 'Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome', *Nature genetics*, 39(3), pp. 311–318. doi: 10.1038/ng1966.

Horikawa, Y. *et al.* (1997) 'Mutation in hepatocyte nuclear factor-1 beta gene (TCF2) associated with MODY.', *Nature genetics*. United States, pp. 384–385. doi: 10.1038/ng1297-384.

Hsu, P. D., Lander, E. S. and Zhang, F. (2014) 'Development and applications of CRISPR-Cas9 for genome engineering', *Cell*. Elsevier, 157(6), pp. 1262–1278. doi: 10.1016/j.cell.2014.05.010.

Hypponen, E. *et al.* (2001) 'Intake of vitamin D and risk of type 1 diabetes: a birth-cohort study.', *Lancet (London, England)*. England, 358(9292), pp. 1500–1503. doi: 10.1016/S0140-6736(01)06580-1.

Ionescu-Tirgoviste, C. *et al.* (2015) 'A 3D map of the islet routes throughout the healthy human pancreas', *Scientific Reports*, 5(1), p. 14634. doi: 10.1038/srep14634.

Jennings, R. E. *et al.* (2013) 'Development of the Human Pancreas From Foregut to Endocrine Commitment TL - 62', *Diabetes*, 62 VN-r(10), pp. 3514–3522. doi: 10.2337/db12-1479.

Jennings, R. E. *et al.* (2015a) 'Human pancreas development.', *Development (Cambridge, England)*, 142(18), pp. 3126–37. doi: 10.1242/dev.120063.

Jennings, R. E. *et al.* (2015b) 'Human pancreas development.', *Development (Cambridge, England)*, 142(18), pp. 3126–37. doi: 10.1242/dev.120063.

Jennings, R. E. *et al.* (2017) 'Laser Capture and Deep Sequencing Reveals the Transcriptomic Programmes Regulating the Onset of Pancreas and Liver Differentiation in Human Embryos', *Stem Cell Reports*. ElsevierCompany., 9, pp. 4–11. doi: 10.1016/J.STEMCR.2017.09.018.

Jeon, J. *et al.* (2009) 'Endocrine cell clustering during human pancreas development.', *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*. United States, 57(9), pp. 811–824. doi: 10.1369/jhc.2009.953307.

Jiang, J. *et al.* (2007) 'Generation of insulin-producing islet-like clusters from human embryonic stem cells.', *Stem cells (Dayton, Ohio)*. United States, 25(8), pp. 1940–1953. doi: 10.1634/stemcells.2006-0761.

Jin, K. and Xiang, M. (2019) 'Transcription factor Ptf1a in development, diseases and reprogramming.', *Cellular and molecular life sciences : CMLS*. Switzerland, 76(5), pp. 921–940. doi: 10.1007/s00018-018-2972-z.

Kanai-Azuma, M. *et al.* (2002) 'Depletion of definitive gut endoderm in Sox17-null mutant mice.', *Development (Cambridge, England)*. England, 129(10), pp. 2367–2379.

Katsarou, A. *et al.* (2017) 'Type 1 diabetes mellitus.', *Nature reviews. Disease primers*. England, 3, p. 17016. doi: 10.1038/nrdp.2017.16.

Kawaguchi, Y. *et al.* (2002) 'The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors.', *Nature genetics*. United States, 32(1), pp. 128–134. doi: 10.1038/ng959.

Kim, S. K. and MacDonald, R. J. (2002) 'Signaling and transcriptional control of pancreatic organogenesis.', *Current opinion in genetics & development*. England, 12(5), pp. 540–547. doi: 10.1016/s0959-437x(02)00338-6.

Kim, S. K. and Melton, D. A. (1998) 'Pancreas development is promoted by cyclopamine, a hedgehog signaling inhibitor.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 95(22), pp. 13036–13041. doi: 10.1073/pnas.95.22.13036.

Kim, W. *et al.* (2010) 'Notch signaling in pancreatic endocrine cell and diabetes', *Biochemical and biophysical research communications*. 2009/12/24, 392(3), pp. 247–251. doi: 10.1016/j.bbrc.2009.12.115.

Kondrashova, A. *et al.* (2013) 'The "Hygiene hypothesis" and the sharp gradient in the incidence of autoimmune and allergic diseases between Russian Karelia and Finland.', *APMIS : acta pathologica, microbiologica, et immunologica Scandinavica*. Denmark, 121(6), pp. 478–493. doi: 10.1111/apm.12023.

Kordowich, S. *et al.* (2012) 'Generation of animals allowing the conditional inactivation of the Pax4 gene', *Transgenic research*. 2012/06/21. Springer Netherlands, 21(6), pp. 1215–1220. doi: 10.1007/s11248-012-9624-0.

Koutsourakis, M. *et al.* (1999) 'The transcription factor GATA6 is essential for early extraembryonic development.', *Development (Cambridge, England)*, 126(9), pp. 723–32. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10383242.

Krapp, A. *et al.* (1998) 'The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas.', *Genes & development*. United States, 12(23), pp. 3752–3763. doi: 10.1101/gad.12.23.3752.

Kroon, E. *et al.* (2008) 'Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo.', *Nature biotechnology*. United States, 26(4), pp. 443–452. doi: 10.1038/nbt1393.

Kuang, J. *et al.* (2014) 'Identification of insulin as a novel retinoic acid receptor-related orphan receptor α target gene', *FEBS Letters*. Federation of European Biochemical Societies, 588(6), pp. 1071–1079. doi: 10.1016/j.febslet.2014.02.029.

Kubo, A. *et al.* (2004) 'Development of definitive endoderm from embryonic stem cells in culture.', *Development (Cambridge, England)*. England, 131(7), pp. 1651–1662. doi:

10.1242/dev.01044.

Lammert, E., Cleaver, O. and Melton, D. (2001) 'Induction of pancreatic differentiation by signals from blood vessels.', *Science (New York, N.Y.)*. United States, 294(5542), pp. 564–567. doi: 10.1126/science.1064344.

Lango Allen, H. *et al.* (2012) 'GATA6 haploinsufficiency causes pancreatic agenesis in humans.', *Nature genetics*. Nature Publishing Group, 44(1), pp. 20–2. doi: 10.1038/ng.1035.

Lau, P. *et al.* (2011) 'Homozygous staggerer (sg/sg) mice display improved insulin sensitivity and enhanced glucose uptake in skeletal muscle.', *Diabetologia*. Germany, 54(5), pp. 1169–1180. doi: 10.1007/s00125-011-2046-3.

Lee, C. S. *et al.* (2002) 'Foxa2 Controls Pdx1 Gene Expression in Pancreatic β-Cells In Vivo', *Diabetes*, 51(8), pp. 2546 LP – 2551. doi: 10.2337/diabetes.51.8.2546.

Lee, K. *et al.* (2019) 'FOXA2 Is Required for Enhancer Priming during Pancreatic Differentiation.', *Cell reports*. United States, 28(2), pp. 382-393.e7. doi: 10.1016/j.celrep.2019.06.034.

Liao, C.-M. *et al.* (2018) 'GATA6 suppression enhances lung specification from human pluripotent stem cells', *The Journal of Clinical Investigation*. The American Society for Clinical Investigation, 128(7), pp. 2944–2950. doi: 10.1172/JCI96539.

Lin, C. *et al.* (2017) 'Development and validation of an ultra-performance liquid chromatography-tandem mass spectrometry method for quantification of SR1001, an inverse agonist of retinoid-related orphan receptors, and its application to pharmacokinetic studies in streptozotoci', *Journal of pharmaceutical and biomedical analysis*. England, 143, pp. 94–100. doi: 10.1016/j.jpba.2017.05.022.

Liu, C. *et al.* (2002) 'GATA-6 and thyroid transcription factor-1 directly interact and regulate surfactant protein-C gene expression.', *The Journal of biological chemistry*. United States, 277(6), pp. 4519–4525. doi: 10.1074/jbc.M107585200.

Loh, K. M. *et al.* (2014) 'Efficient endoderm induction from human pluripotent stem cells by logically directing signals controlling lineage bifurcations.', *Cell stem cell*. United States, 14(2), pp. 237–252. doi: 10.1016/j.stem.2013.12.007.

Lorberbaum, D. S. and Sussel, L. (2017) 'Gotta Have GATA for Human Pancreas Development.', *Cell stem cell*. United States, 20(5), pp. 577–579. doi: 10.1016/j.stem.2017.04.004.

Lyttle, B. M. *et al.* (2008) 'Transcription factor expression in the developing human fetal endocrine pancreas.', *Diabetologia*. Germany, 51(7), pp. 1169–1180. doi: 10.1007/s00125-008-1006-z.

MacDonald, M. J. *et al.* (2011) 'Differences between human and rodent pancreatic islets: low pyruvate carboxylase, atp citrate lyase, and pyruvate carboxylation and high glucose-stimulated acetoacetate in human pancreatic islets.', *The Journal of biological chemistry*. United States, 286(21), pp. 18383–18396. doi: 10.1074/jbc.M111.241182.

Marchetti, P. *et al.* (2017) 'Pancreatic Beta Cell Identity in Humans and the Role of Type 2 Diabetes.', *Frontiers in cell and developmental biology*. Switzerland, 5, p. 55. doi: 10.3389/fcell.2017.00055.

Martinelli, P. *et al.* (2013) 'Gata6 is required for complete acinar differentiation and maintenance of the exocrine pancreas in adult mice.', *Gut*. England, 62(10), pp. 1481–1488. doi: 10.1136/gutjnl-2012-303328.

Mcclellan, J. and King, M. (2010) 'Genetic Heterogeneity in Human Disease', *Cell*, 141(2), pp. 210–217. doi: 10.1016/j.cell.2010.03.032.

McCracken, K. W. *et al.* (2017) 'Wnt/β-catenin promotes gastric fundus specification in mice and humans', *Nature*. Nature Publishing Group, 541(7636), pp. 182–187. doi: 10.1038/nature21021.

Micallef, S. J. *et al.* (2012) 'INS(GFP/w) human embryonic stem cells facilitate isolation of in vitro derived insulin-producing cells.', *Diabetologia*. Germany, 55(3), pp. 694–706. doi: 10.1007/s00125-011-2379-y.

Milne, R. L. and Antoniou, A. C. (2011) 'Genetic modifiers of cancer risk for BRCA1 and BRCA2 mutation carriers.', *Annals of oncology : official journal of the European Society for Medical Oncology*. England, 22 Suppl 1, pp. i11-7. doi: 10.1093/annonc/mdq660.

Molkentin, J. D. (2000) 'The zinc finger-containing transcription factors GATA-4, -5, and -6. Ubiquitously expressed regulators of tissue-specific gene expression.', *The Journal of biological chemistry*, 275(50), pp. 38949–52. doi: 10.1074/jbc.R000029200.

Morrisey, E. E. *et al.* (1996) 'GATA-6: a zinc finger transcription factor that is expressed in multiple cell lineages derived from lateral mesoderm.', *Developmental biology*. United States, 177(1), pp. 309–322. doi: 10.1006/dbio.1996.0165.

Morrisey, E. E. *et al.* (1998) 'GATA6 regulates HNF4 and is required for differentiation of visceral endoderm in the mouse embryo.', *Genes & development*, 12(22), pp. 3579–90. Available at: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=317242&tool=pmcentrez&rendertype=abstract.

Mühlbauer, E. *et al.* (2013) 'Differential and day-time dependent expression of nuclear receptors RORα, RORβ, RORγ and RXRα in the rodent pancreas and islet', *Molecular and Cellular Endocrinology*, 365(2), pp. 129–138. doi: https://doi.org/10.1016/j.mce.2012.10.001.

Murtaugh, L. C. (2007) 'Pancreas and beta-cell development: from the actual to the possible', *Development*, 134(3), pp. 427 LP – 438. doi: 10.1242/dev.02770.

Musunuru, K. (2013) 'Genome editing of human pluripotent stem cells to generate human cellular disease models', *Disease models & mechanisms*, 6(4), pp. 896–904. doi: 10.1242/dmm.012054.

Nair, G. G. *et al.* (2019) 'Recapitulating endocrine cell clustering in culture promotes maturation of human stem-cell-derived beta cells.', *Nature cell biology*. England, 21(2), pp. 263–274. doi: 10.1038/s41556-018-0271-4.

Nostro, M Cristina *et al.* (2015) 'Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines.', *Stem cell reports*. United States, 4(4), pp. 591–604. doi: 10.1016/j.stemcr.2015.02.017.

Nostro, M. Cristina *et al.* (2015) 'Efficient generation of NKX6-1+ pancreatic progenitors from multiple human pluripotent stem cell lines', *Stem Cell Reports*. The Authors, 4(4), pp. 591–604. doi: 10.1016/j.stemcr.2015.02.017.

O'Rahilly, R. and Muller, F. (2010) 'Developmental stages in human embryos: revised and new measurements.', *Cells, tissues, organs*. Switzerland, 192(2), pp. 73–84. doi: 10.1159/000289817.

Offield, M. F. *et al.* (1996) 'PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum.', *Development (Cambridge, England)*. England, 122(3), pp. 983–995.

Otter, S. and Lammert, E. (2016) 'Exciting Times for Pancreatic Islets: Glutamate Signaling in Endocrine Cells', *Trends in Endocrinology and Metabolism*. Elsevier Ltd, 27(3), pp. 177–188. doi: 10.1016/j.tem.2015.12.004.

Pagliuca, F. W. *et al.* (2014) 'Generation of functional human pancreatic beta cells in vitro.', *Cell*. United States, 159(2), pp. 428–439. doi: 10.1016/j.cell.2014.09.040.

Pan, F. C. *et al.* (2013) 'Spatiotemporal patterns of multipotentiality in Ptf1a-expressing cells during pancreas organogenesis and injury-induced facultative restoration.', *Development (Cambridge, England)*. England, 140(4), pp. 751–764. doi: 10.1242/dev.090159.

Pan, F. C. and Wright, C. (2011) 'Pancreas organogenesis: From bud to plexus to gland', *Developmental Dynamics*, 240(3), pp. 530–565. doi: 10.1002/dvdy.22584.

Pandol, S. J. (2010) *The Exocrine Pancreas*. San Rafael (CA). doi: 10.4199/C00026ED1V01Y201102ISP014.

Park, I.-H. *et al.* (2008) 'Disease-specific induced pluripotent stem cells', *Cell*. 2008/08/07, 134(5), pp. 877–886. doi: 10.1016/j.cell.2008.07.041.

Park, J. *et al.* (2011) 'Distribution of allele frequencies and effect sizes and their interrelationships for common genetic susceptibility variants', *Proceedings of the National Academy of Sciences of the United States of America*, 108(44), pp. 18026–18031. doi: 10.1073/pnas.1114759108.

Parkin, C. A. and Ingham, P. W. (2008) 'The adventures of Sonic Hedgehog in development and repair. I. Hedgehog signaling in gastrointestinal development and disease.', *American journal of physiology. Gastrointestinal and liver physiology*. United States, 294(2), pp. G363-7. doi: 10.1152/ajpgi.00457.2007.

Pashos, E. E. *et al.* (2017) 'Large , Diverse Population Cohorts of hiPSCs and Derived Hepatocyte-like Cells Reveal Functional Genetic Variation at Blood Lipid-Associated Loci Resource Large , Diverse Population Cohorts of hiPSCs and Derived Hepatocyte-like Cells Reveal Functional Gen', *Stem Cell*. Elsevier Inc., 20(4), pp. 558-570.e10. doi: 10.1016/j.stem.2017.03.017.

Penvose, A. *et al.* (2019) 'Comprehensive study of nuclear receptor DNA binding provides a revised framework for understanding receptor specificity.', *Nature communications*. England, 10(1), p. 2514. doi: 10.1038/s41467-019-10264-3.

Pierzynowski, S. G. *et al.* (2018) 'Glucose homeostasis dependency on acini–islet–acinar (AIA) axis communication: a new possible pathophysiological hypothesis regarding diabetes mellitus', *Nutrition & Diabetes*, 8(1), p. 55. doi: 10.1038/s41387-018-0062-9.

Piper, K. *et al.* (2004) 'Beta cell differentiation during early human pancreas development.', *The Journal of endocrinology*. England, 181(1), pp. 11–23. doi: 10.1677/joe.0.1810011.

Quack, M. and Carlberg, C. (2001) 'Single thyroid hormone receptor monomers are competent for co-activator-mediated transactivation.', *The Biochemical journal*. England, 360(Pt 2), pp. 387–393. doi: 10.1042/0264-6021:3600387.

Rezania, A. *et al.* (2014) 'Reversal of diabetes with insulin-producing cells derived in vitro from human pluripotent stem cells', *Nat Biotechnol*, 32(11), pp. 1121–1133. doi: 10.1038/nbt.3033.

Rhodes, C. J. (2005) 'Type 2 diabetes-a matter of beta-cell life and death?', *Science (New York, N.Y.)*. United States, 307(5708), pp. 380–384. doi: 10.1126/science.1104345.

Ripke, S. *et al.* (2013) 'Genome-wide association analysis identifies 13 new risk loci for schizophrenia.', *Nature genetics*. United States, 45(10), pp. 1150–1159. doi: 10.1038/ng.2742.

Rodriguez-Segui, S., Akerman, I. and Ferrer, J. (2012) 'GATA believe it: new essential regulators of pancreas development.', *The Journal of clinical investigation*. United States, 122(10), pp. 3469–3471. doi: 10.1172/JCI65751.

Rodríguez-Seguí, S., Akerman, I. and Ferrer, J. (2012) 'GATA believe it: new essential regulators of pancreas development.', *The Journal of clinical investigation*, 122(10), pp. 3469–71. doi: 10.1172/JCI65751.

Roussos, P. *et al.* (2014) 'A role for noncoding variation in schizophrenia.', *Cell reports*. United States, 9(4), pp. 1417–1429. doi: 10.1016/j.celrep.2014.10.015.

Rubio-Cabezas, O. *et al.* (2014) 'ISPAD Clinical Practice Consensus Guidelines 2014. The diagnosis and management of monogenic diabetes in children and adolescents.', *Pediatric diabetes*. Denmark, 15 Suppl 2, pp. 47–64. doi: 10.1111/pedi.12192.

Russ, H. A. *et al.* (2015) 'Controlled induction of human pancreatic progenitors produces functional beta-like cells in vitro.', *The EMBO journal*. England, 34(13), pp. 1759–1772. doi: 10.15252/embj.201591058.

Salisbury, R. J. *et al.* (2014) 'The window period of NEUROGENIN3 during human gestation', *Islets*. Taylor & Francis, 6(3), pp. e954436–e954436. doi: 10.4161/19382014.2014.954436.

Saltiel, A. R. and Kahn, C. R. (2001) 'Insulin signalling and the regulation of glucose and lipid metabolism.', *Nature*. England, 414(6865), pp. 799–806. doi: 10.1038/414799a.

Sander, M. *et al.* (2000) 'Homeobox gene Nkx6.1 lies downstream of Nkx2.2 in the major pathway of beta-cell formation in the pancreas.', *Development (Cambridge, England)*. England, 127(24), pp. 5533–5540.

Sandip, S. *et al.* (2016) 'Dorsal pancreatic agenesis: description of CT signs', *BMJ case reports*. BMJ Publishing Group, 2016, p. bcr2015210093. doi: 10.1136/bcr-2015-210093.

Sareen, D. *et al.* (2013) 'Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion.', *Science translational medicine*. United States, 5(208), p. 208ra149. doi: 10.1126/scitranslmed.3007529.

Sartori, D. J. *et al.* (2014) 'GATA factors promote ER integrity and β-cell survival and contribute to type 1 diabetes risk.', *Molecular endocrinology (Baltimore, Md.)*, 28(1), pp. 28–39. doi: 10.1210/me.2013-1265.

Seino, S., Shibasaki, T. and Minami, K. (2011) 'Dynamics of insulin secretion and the clinical implications for obesity and diabetes.', *The Journal of clinical investigation*. United States, 121(6), pp. 2118–2125. doi: 10.1172/JCI45680.

Sellick, G. S., Barker, K. T., Stolte-Dijkstra, I., *et al.* (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis.', *Nature genetics*. United States, 36(12), pp. 1301–1305. doi: 10.1038/ng1475.

Sellick, G. S., Barker, K. T., Stolte-dijkstra, I., *et al.* (2004) 'Mutations in PTF1A cause pancreatic and cerebellar agenesis', *Nature genetics*, 36(12), pp. 1301–1305. doi: 10.1038/ng1475.

Seymour, P. A. (2014) 'Sox9: a master regulator of the pancreatic program', *The review of diabetic studies : RDS*. 2014/05/10. SBDR - Society for Biomedical Diabetes Research, 11(1), pp. 51–83. doi: 10.1900/RDS.2014.11.51.

Shalem, O., Sanjana, N. E. and Zhang, F. (2015) 'High-throughput functional genomics using CRISPR-Cas9.', *Nature reviews. Genetics*. England, 16(5), pp. 299–311. doi: 10.1038/nrg3899.

Shaw-smith, C. *et al.* (2014) 'GATA4 Mutations Are a Cause of Neonatal and Childhood- Onset Diabetes', 63(January), pp. 2888–2894. doi: 10.2337/db14-0061.

Shaw-Smith, C. *et al.* (2014) 'GATA4 mutations are a cause of neonatal and childhood-onset diabetes', *Diabetes*, 63(8), pp. 2888–2894. doi: 10.2337/db14-0061.

Sheets, T. P. *et al.* (2018) 'Targeted Mutation of NGN3 Gene Disrupts Pancreatic Endocrine Cell Development in Pigs', *Scientific Reports*, 8(1), p. 3582. doi: 10.1038/s41598-018-22050-0.

Shi, Z. D. *et al.* (2017) 'Genome Editing in hPSCs Reveals GATA6 Haploinsufficiency and a Genetic Interaction with GATA4 in Human Pancreatic Development', *Cell Stem Cell*. Elsevier Inc., 20(5), pp. 675-688.e6. doi: 10.1016/j.stem.2017.01.001.

Shih, H. P., Wang, A. and Sander, M. (2013) 'Pancreas organogenesis: from lineage determination to morphogenesis.', *Annual review of cell and developmental biology*, 29, pp. 81–105. doi: 10.1146/annurev-cellbio-101512-122405.

Da Silva Xavier, G. (2018) 'The Cells of the Islets of Langerhans', *Journal of clinical medicine*. MDPI, 7(3), p. 54. doi: 10.3390/jcm7030054.

Sinner, D. *et al.* (2004) 'Sox17 and beta-catenin cooperate to regulate the transcription of endodermal genes.', *Development (Cambridge, England)*. England, 131(13), pp. 3069–3080. doi: 10.1242/dev.01176.

Soldner, F. *et al.* (2016) 'Parkinson-associated risk variant in distal enhancer of α-synuclein modulates target gene expression', *Nature*. Nature Publishing Group, 533(7601), pp. 95–99. doi: 10.1038/nature17939.

Solt, L. A. and Burris, T. P. (2012) 'Action of RORs and their ligands in (patho) physiology', *Trends in Endocrinology & Metabolism*. Elsevier Ltd, 23(12), pp. 573–581. doi: 10.1016/j.tem.2012.05.012.

Spence, J. R. *et al.* (2009) 'Sox17 regulates organ lineage segregation of ventral foregut progenitor cells.', *Developmental cell*. United States, 17(1), pp. 62–74. doi: 10.1016/j.devcel.2009.05.012.

Stanescu, D. E. *et al.* (2015) 'A novel mutation in GATA6 causes pancreatic agenesis', *Pediatric Diabetes*, 16(1), pp. 67–70. doi: 10.1111/pedi.12111.

Steck, A. K. and Winter, W. E. (2011) 'Review on monogenic diabetes.', *Current opinion in endocrinology, diabetes, and obesity*. England, 18(4), pp. 252–258. doi: 10.1097/MED.0b013e3283488275.

Stoffers, D. A. *et al.* (1997) 'Pancreatic agenesis attributable to a single nucleotide deletion in the human IPF1 gene coding sequence', *Nature Genetics*. Nature Publishing Group, 15, p. 106. Available at: https://doi.org/10.1038/ng0197-106.

Sullivan, S. K. *et al.* (2014) 'High-level transgene expression in induced pluripotent stem cell-derived megakaryocytes: correction of Glanzmann thrombasthenia.', *Blood*. United States, 123(5), pp. 753–757. doi: 10.1182/blood-2013-10-530725.

Suzuki, E. *et al.* (1996) 'The human GATA-6 gene: structure, chromosomal location, and regulation of expression by tissue-specific and mitogen-responsive signals.', *Genomics*, 38(3), pp. 283–290. doi: 10.1006/geno.1996.0630.

Suzuki, S. *et al.* (2014) 'A case of pancreatic agenesis and congenital heart defects with a novel GATA6 nonsense mutation: Evidence of haploinsufficiency due to nonsense-mediated mRNA decay', *American Journal of Medical Genetics, Part A*, 164(2), pp. 476–479. doi: 10.1002/ajmg.a.36275.

Takahashi, K. *et al.* (2007) 'Induction of pluripotent stem cells from adult human fibroblasts by defined factors.', *Cell*. United States, 131(5), pp. 861–872. doi: 10.1016/j.cell.2007.11.019.

Takahashi, K. and Yamanaka, S. (2006) 'Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.', *Cell*. United States, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Tam, P. P. and Behringer, R. R. (1997) 'Mouse gastrulation: the formation of a mammalian body plan.', *Mechanisms of development*. Ireland, 68(1–2), pp. 3–25. doi: 10.1016/s0925-4773(97)00123-8.

Tattersall, R. B. (1974) 'Mild familial diabetes with dominant inheritance.', *The Quarterly journal of medicine*. England, 43(170), pp. 339–357.

Teo, A. K. K. *et al.* (2015) 'PDX1 binds and repreßes hepatic genes to ensure robust pancreatic commitment in differentiating human embryonic stem cells', *Stem Cell Reports*. The Authors, 4(4), pp. 578–590. doi: 10.1016/j.stemcr.2015.02.015.

Thomson, J. A. *et al.* (1998) 'Embryonic stem cell lines derived from human blastocysts.', *Science (New York, N.Y.)*. United States, 282(5391), pp. 1145–1147. doi: 10.1126/science.282.5391.1145.

Thurman, R. E. *et al.* (2012) 'The accessible chromatin landscape of the human genome', *Nature*. Nature Publishing Group, 488(7414), pp. 75–82. doi: 10.1038/nature11232.

Tiyaboonchai, A. *et al.* (2017) 'GATA6 Plays an Important Role in the Induction of Human Definitive Endoderm, Development of the Pancreas, and Functionality of Pancreatic β Cells', *Stem Cell Reports*, 8, pp. 589–604. doi: 10.1016/j.stemcr.2016.12.026.

Trynka, G. *et al.* (2013) 'Chromatin marks identify critical cell types for fine mapping complex trait variants.', *Nature genetics*. United States, 45(2), pp. 124–130. doi: 10.1038/ng.2504.

Velazco-Cruz, L. *et al.* (2019) 'Acquisition of Dynamic Function in Human Stem Cell-Derived beta Cells.', *Stem cell reports*. United States, 12(2), pp. 351–365. doi: 10.1016/j.stemcr.2018.12.012.

Veres, A. *et al.* (2019) 'Charting cellular identity during human in vitro beta-cell differentiation.', *Nature*. England, 569(7756), pp. 368–373. doi: 10.1038/s41586-019-1168-5.

Viger, R. S. *et al.* (2008) 'Role of the GATA Family of Transcription Factors in Endocrine Development, Function, and Disease', *Molecular Endocrinology*, 22(4), pp. 781–798. doi: 10.1210/me.2007-0513.

Villasenor, A. *et al.* (2010) 'Epithelial dynamics of pancreatic branching morphogenesis.', *Development (Cambridge, England)*, 137(24), pp. 4295–305. doi: 10.1242/dev.052993.

Visscher, P. M. *et al.* (2017) '10 Years of GWAS Discovery : Biology , Function , and Translation', *The American Journal of Human Genetics*. ElsevierCompany., 101(1), pp. 5–22. doi: 10.1016/j.ajhg.2017.06.005.

Wang, D. *et al.* (2018) 'Interactions Between Regulatory Variants in CYP7A1 (Cholesterol 7alpha-Hydroxylase) Promoter and Enhancer Regions Regulate CYP7A1 Expression.', *Circulation. Genomic and precision medicine*. United States, 11(10), p. e002082. doi: 10.1161/CIRCGEN.118.002082.

Wareham, N. J. and Herman, W. H. (2016) 'The Clinical and Public Health Challenges of Diabetes Prevention: A Search for Sustainable Solutions.', *PLoS medicine*. United States, p. e1002097. doi: 10.1371/journal.pmed.1002097.

Warren, C. R. *et al.* (2017) 'Induced Pluripotent Stem Cell Differentiation Enables Functional Validation of GWAS Variants in Metabolic Disease.', *Cell stem cell*. United States, 20(4), pp. 547-557.e7. doi: 10.1016/j.stem.2017.01.010.

Weedon, M. N., Cebola, Inês, *et al.* (2014) 'Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis.', *Nature genetics*. Nature Publishing Group, 46(1), pp. 61–4. doi: 10.1038/ng.2826.

Weedon, M. N., Cebola, Ines, *et al.* (2014) 'Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis', 46(1), pp. 61–64. doi: 10.1038/ng.2826.Recessive.

Wei, W.-H., Hemani, G. and Haley, C. S. (2014) 'Detecting epistasis in human complex traits.', *Nature reviews. Genetics*. England, 15(11), pp. 722–733. doi: 10.1038/nrg3747.

Weintraub, H. *et al.* (1989) 'Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD.', *Proceedings of the National Academy of Sciences of the United States of America*. United States, 86(14), pp. 5434–5438. doi: 10.1073/pnas.86.14.5434.

Weirauch, M. T. *et al.* (2014) 'Determination and inference of eukaryotic transcription factor sequence specificity.', *Cell*. United States, 158(6), pp. 1431–1443. doi: 10.1016/j.cell.2014.08.009.

Whiting, D. R. *et al.* (2011) 'IDF diabetes atlas: global estimates of the prevalence of diabetes for 2011 and 2030.', *Diabetes research and clinical practice*. Ireland, 94(3), pp. 311–321. doi: 10.1016/j.diabres.2011.10.029.

Willet, S. G. and Mills, J. C. (2016) 'Stomach Organ and Cell Lineage Differentiation: From Embryogenesis to Adult Homeostasis', *Cellular and Molecular Gastroenterology and Hepatology*, 2(5), pp. 546–559. doi: https://doi.org/10.1016/j.jcmgh.2016.05.006.

Xuan, S. *et al.* (2012) 'Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis', *Journal of Clinical Investigation*, 122(10), pp. 3516–3528. doi: 10.1172/JCI63352.

Xuan, S. and Sussel, L. (2016) 'GATA4 and GATA6 regulate pancreatic endoderm identity through inhibition of hedgehog signaling.', *Development (Cambridge, England)*, 143(5), pp. 780–6. doi: 10.1242/dev.127217.

Yamanaka, S. (2012) 'Induced pluripotent stem cells: past, present, and future.', *Cell stem cell*. United States, 10(6), pp. 678–684. doi: 10.1016/j.stem.2012.05.005.

Yau, D. *et al.* (2017) 'Case report : maternal mosaicism resulting in inheritance of a novel GATA6 mutation causing pancreatic agenesis and neonatal diabetes mellitus', *Diagnostic Pathology*. Diagnostic Pathology, pp. 1–7. doi: 10.1186/s13000-016-0592-1.

Yorifuji, T. *et al.* (2012) 'Dominantly inherited diabetes mellitus caused by GATA6 haploinsufficiency: variable intrafamilial presentation.', *Journal of medical genetics*, 49(10), pp. 642–3. doi: 10.1136/jmedgenet-2012-101161.

Yu, L. *et al.* (2014) 'Whole exome sequencing identifies de novo mutations in GATA6 associated with congenital diaphragmatic hernia.', *Journal of medical genetics*, 51(3), pp. 197–202. doi: 10.1136/jmedgenet-2013-101989.

Zhang, D. *et al.* (2009) 'Highly efficient differentiation of human ES cells and iPS cells into mature pancreatic insulin-producing cells.', *Cell research*. England, 19(4), pp. 429–438. doi: 10.1038/cr.2009.28.

Zhang, F. and Lupski, J. R. (2015) 'Non-coding genetic variants in human disease', *Human molecular genetics*, 24(R1), pp. R102-10. doi: 10.1093/hmg/ddv259.

Zhang, Yuwei *et al.* (2016) 'Genetic Variants of Retinoic Acid Receptor-Related Orphan Receptor Alpha Determine Susceptibility to Type 2 Diabetes Mellitus in Han Chinese', *Genes*. MDPI, 7(8), p. 54. doi: 10.3390/genes7080054.

Zhao, R. *et al.* (2005) 'GATA6 Is Essential for Embryonic Development of the Liver but Dispensable for Early Heart Formation', *Molecular and Cellular Biology*, 25(7), pp. 2622 LP – 2631. doi: 10.1128/MCB.25.7.2622-2631.2005.

Zheng, Y., Ley, S. H. and Hu, F. B. (2018) 'Global aetiology and epidemiology of type 2 diabetes mellitus and its complications.', *Nature reviews. Endocrinology*. England, 14(2), pp. 88–98. doi: 10.1038/nrendo.2017.151.